

# The conscious awareness of visual space: A tripartite encoding model.

Dhanraj Vishwanath<sup>1</sup>

<sup>1</sup>School of Psychology & Neuroscience

University of St. Andrews

\*In Press: Psychology of Consciousness: Theory Methods and Practice

© 2021, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/cns0000280

## Author Note

Correspondence concerning this article should be addressed to Dhanraj Vishwanath

School of Psychology & Neuroscience, University of St. Andrews, St. Andrews, Fife,  
Scotland UK KY16 9JP

Dv10@st-andrews.ac.uk

## **Abstract**

The prevailing model of 3D vision proposes that the visual system recovers a single objective and internally consistent representation of physical 3D space based on a process of ideal-observer probabilistic inference. A significant challenge for this model has been in explaining the contents of our subjective awareness of visual space. Here I argue that integrating phenomenological observations, empirical data, evolutionary logic and neurophysiological evidence leads to the conjecture that the human conscious awareness of visual space is underwritten by multiple, sometimes mutually inconsistent, spatial encodings. By assessing four primary competencies in the conscious awareness of space, three major types of spatial encodings are conjectured. Among the most primitive of these is proposed to support the competency of the conscious awareness of distance at an ambulatory scale (operationally defined as egocentric distance) and is hypothesised to involve temporal archicortex regions. The second is proposed to support the competency of awareness of object layout and 3D shape without scale (operationally, relative depth), likely instantiated in the ventral visual stream of the neocortex. This encoding is hypothesised to have evolved from more primitive encodings that provide a depth-ordered segmentation of the visual field. The third encoding is proposed to support the competency of fine-grained awareness of intra- and inter-object spatial separation in near space (operationally, scaled or absolute depth) and instantiated in the dorsal visual stream. This encoding is conjectured to underlie the phenomenology of object solidity, spatial separation, tangibility and object realness that is often referred to as stereopsis. The combined effect of the first and third competencies (ambulatory distance and near-space scaled spatial separation) is conjectured to contribute to the feeling of spatial immersion and presence.

## Introduction

When we look out into the external world, we have conscious visual awareness of a 3-dimensional space inhabited at various locations by 3-dimensional objects. The most widely accepted textbook explanation for this ability is that the brain reconstructs a “representation” of the physical 3D scene, and that it is this representation that we have conscious awareness of (e.g., Sekuler & Blake, 2006; Palmer, 1999). According to this view, the reconstruction is based on spatial estimates arising from various sources of sensory information (referred to as depth cues) either in the retinal image (e.g., binocular disparity, perspective convergence) or from extra-retinal sources (e.g., ocular vergence, declination from eye level). The underlying assumption is that this reconstruction, under normal operating conditions, leads to the inference of a single objective and veridical (i.e., accurate) representation of 3D space. The objectivity assumption is based on the belief that the visual representation is descriptively isomorphic to physical space, which, according to classical physics, is best described as a 3-dimensional Euclidean metric space. The veridicality claim is that the values of spatial attributes constituted in the visual representation (point location coordinates, distance, depth, slant, curvature, etc.) do not deviate in any systematic quantitative way from the actual physical spatial values, except in special cases such as visual illusions. This claim is incorporated in the “unbiasedness” assumption in the most widely accepted version of the prevailing model of 3D vision: the probabilistic-cue-integration model of depth perception (Landy et al., 1995; Landy et al., 2011). Early versions of this model claims that depth cues yield gaussian-distributed estimates of spatial parameters where the mean of the distribution is unbiased (i.e., matches the physical ground truth). Later versions have considered non-Gaussian distributions (e.g., Knill, 2007; Saunders & Knill, 2001) where instead of the mean, it is the distribution itself that is assumed to be encoded accurately by the visual system.

In summary, the prevailing approach to 3D space perception suggests that we are consciously aware of a visual space that is singular, more or less accurate, internally consistent and constitutes an *objective* representation that is isomorphic to a physical description of the scene.

Closer inspection of our conscious awareness of visual space, however, reveals challenges to this model and its assumptions. Consider the ubiquitous example of a correct-perspective picture or photograph of a real scene. The picture conveys an impression of a 3D space and

objects which, in many important ways, is the same as the impression obtained viewing the real scene. However, there are differences. Compared to viewing real scenes, pictorial objects lack the phenomenological impression of solidity and tangibility characteristic of real objects and the vivid impression of separation (negative space) between objects in depth. Both these qualitative attributes contribute to our sensation of spatial immersion/presence and object realness (Vishwanath, 2010, 2014). In viewing pictures, there is also a visual duality in that we perceive both a flat picture surface located in real space and a virtual pictorial 3D space (Niederée & Heyer, 2003) that appears largely undistorted in comparison to the real scene which it depicts. Finally, while in real-world viewing we generally have a clear impression of the absolute scale of objects in a scene, scale in picture (absent familiar sized objects) is ambiguous<sup>1</sup> (Vishwanath, 2010). These observations regarding pictorial images challenge the assumptions of the prevailing model of 3D perception outlined above. The mere fact that we have conscious awareness of 3D objects arranged in a virtual 3D space when viewing a flat 2D picture contradicts the assumption of veridical reconstruction and objectivity. They also challenge the assumption that a single representation underlies conscious awareness of space.

Consider a second example of the difference in the phenomenological impression of spatial separation (negative space) among objects comparing monocular and binocular viewing of a real scene. Closing one eye when viewing a real scene appears to not appreciably alter the 3D percept: objects still appear to have the same 3D shape and locations with respect to each other, but the phenomenological impression of spatial separation (negative space) between objects appears diminished (Vishwanath, 2014). For example, looking at several objects separated in depth, the relative depth relations within and among objects (3D shape and layout) appears unchanged, and yet, paradoxically, objects appear phenomenologically less separated from each other. A related observation is the difference in the phenomenological impression of spatial separation between objects close to and far away from the observer. For example, two objects separated from each other in depth by a hundred meters viewed from (say) a couple hundred meters away from the observer will appear to have a weaker phenomenological impression of spatial separation than that observed between two objects viewed from a meter away and separated by only 30cm (Vishwanath, 2014). These observations again challenge the internal consistency and singular representation assumption of prevailing models.

---

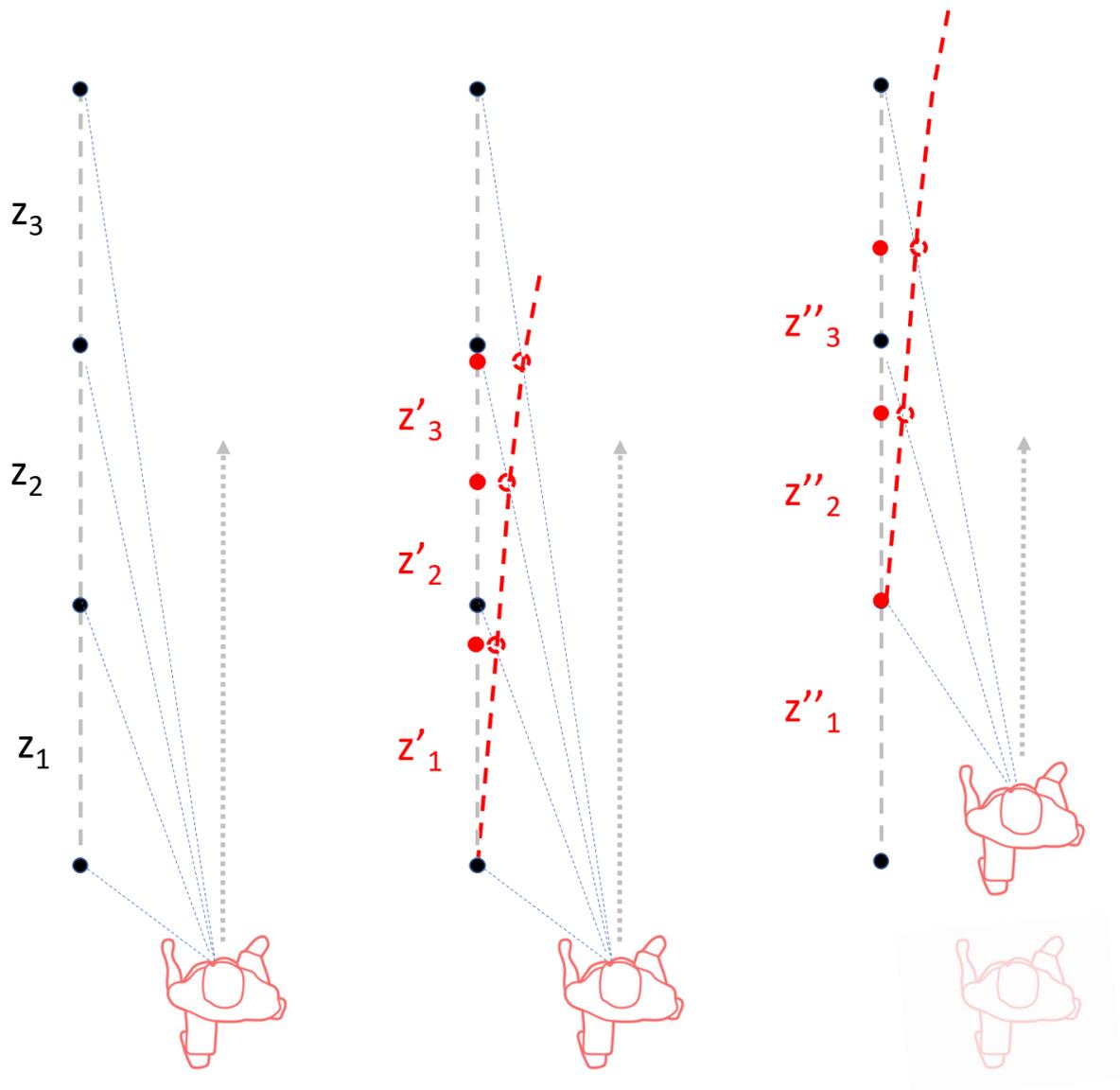
<sup>1</sup> See figure 4b

As a third example, consider standing on a straight path flanked on one side by equally spaced lampposts (or markings on a wall) where the nearest post/marking is about 3 to 5 meters in front of you. Looking at each post/marking individually, one has a clear apprehension or awareness of its distance. However, the relative distances between posts appear unequal, in that, farther distances between posts/markings appear more foreshortened (Fig 1; Erkelens, 2015). This is reinforced by the observation that as you walk along the path, the distance between the nearest two posts/markings appears to expand as you get closer (Fig 2). This phenomenological observation is related to a remarkable empirical finding from studies in blind walking (Loomis et al., 1992; Loomis, Philbeck & Zahorik, 2002). Posts are set at different distances (say, 10 and 12 meters). Subjects are asked to view one of the posts, internalise its perceived distance, turn 90 degrees and walk blindfolded for the length of the remembered distance. This is done for each post separately. Human observers are surprisingly accurate and precise at this task for distances exceeding 20m (Loomis et al., 1992). However, if the observer is asked to perceptually report the distance between two posts in a relative depth task, they significantly underestimate it (Loomis et al., 1992; Loomis et al 2002); particularly under monocular viewing, even though distance estimates from blind walking responses are accurate for monocular viewing as well (Loomis et al, 2002; Ooi & He, 2015). In other words, despite an accurate conscious awareness of the distance to each post, there is a misapprehension of the distances between them (Figure 1). Taken together, these informal and empirical observations suggest a dissociation between the awareness of egocentric distance and the awareness of relative distances (Loomis et al., 2002).



**Figure 1**

Equidistant markings on a planar surface receding in depth. The spatial intervals do not appear equal. Observing an equivalent real scene also produces a similar effect of foreshortening. Empirical results show that observers set the farther separation to be significantly larger than the closer separation for the separations to appear equal in size (Erkelens, 2015; Image adapted from Erkelens, 2015).



**Figure 2**

Left: An observer views equidistant markings on a wall (or posts lying on single imaginary plane). Human observers are relatively accurate in judging the egocentric distances ( $d_1$ ,  $d_2$ ,  $d_3$ ) as demonstrated in blind-walking paradigms. However, the spatial intervals between markings or posts ( $z_1$ ,  $z_2$ ,  $z_3$ ) are typically misperceived (underestimated) as suggested by the red dots in the middle panel, with greater underestimation for farther intervals. Since viewing direction is fixed by retinal visual angles, this implies (based on an assumption of Euclidean space) that the perceived surface defined by the posts should appear slanted or curved away from the true orientation of the post depending on the degree of underestimation. Informal phenomenological observation demonstrates that as the observer approaches the first spatial interval, it will appear to paradoxically expand with no apparent change in the position of the posts or change in the perceived slant of the wall or row of posts.

In summary, the above three cases of phenomenological observations, along with associated empirical results, argue against the assumption that our conscious awareness of space is based on a singular, veridical, internally consistent and objective representation of 3D space that is isomorphic to a physical description of the scene. It suggests the existence of multiple distinct and biased underlying encodings of space, each peculiar to the specific adaptive needs of the agent, resulting in an internally inconsistent subjective awareness of spatial attributes.

While it is generally acknowledged that there are likely different representations underlying visual perception and visual guidance of movement deriving from early findings in neurophysiology (see Mishkin et al., 1983), the mainstream view in vision science has generally been the assumption of a single veridical representation for perception. However, here have been several researchers who have provided empirical results which suggest that the underlying visual representation is not veridical, internally consistent, can be task dependent or implies the existence of more than one representation. For example, depth judgements based on the cue of binocular disparity have been shown to have different biases under different psychophysical tasks suggesting that different depth perception tasks may sample different stages of an underlying representation or different representations (Glennerster et al., 1996; Lappin & Craft, 2000). Several studies on depth representation in personal space (<2m) have demonstrated internal inconsistencies in visual judgements as well as demonstrating that single and multiple depth cues often result in biased (non-veridical) estimates (di Luca et al, 2001; Domini, Caudek & Richman 1998; Domini & Braunstein, 1998; Todd, 2004). Other work examining space perception at ambulatory distances have also shown internal inconsistencies in spatial representation which further suggest the existence of multiple biased representations (Loomis et al., 2002; Svarverud, Gilson & Glennerster, 2012).

However, despite this body of evidence, the current understanding of how the brain recovers a representation of visual space has been dominated by the ideal-observer probabilistic cue integration approach (Landy et al., 2011) since the mid 1990's. Studies that have taken this latter approach have generally not addressed the challenges to the underlying assumption of a single objective and veridical (unbiased) representation of space and have not sought to address phenomena in the conscious awareness of space.

## Visual representation and sensory calibration

A central reason why the prevailing model of 3D vision has struggled to address phenomena in the conscious awareness of space relates to underlying assumptions regarding the *content* of visual representation and how that content is recovered.

The first, albeit implicit assumption, is that the content of visual representation of 3D space is essentially a map or hierarchy of quantitative properties. This might be defined in terms of a first-order depth map of the 3-dimensional spatial coordinates of points in the scene, or a higher order map of geometric values or attributes that compactly summarizes a delimited collection of point coordinates; for example, the orientations of surface normals (surface slant or curvature). The second assumption is that the main goal of understanding 3D space and depth perception is to examine how individual depth cues and their combinations contribute to an accurate recovery of this map of quantitative values (Landy et al., 1995).

These implicit assumptions derive from an approach that casts the understanding of 3D spatial vision in terms of the psychophysical problem of the correlation between quantitative perceptual judgments and the associated physical measures (ground truth). In other words, the problem of 3D vision in the prevailing model is viewed as one of *calibration* (Vishwanath, 2005) and the assumptions derive directly from this view: the objectivity assumption of the presumed isomorphism of visual description with physical reality comes from defining the ground truth for establishing calibration in terms of physical Euclidean space; the notion of veridicality comes for the assumption that the visual system aims for an accurate calibration with ground truth; the idea of visual representation as a collection of quantitative values falls out of the fact that calibration can only be verified by comparing quantities.

The conventional understanding of depth cues and their role in 3D visual representation is also fundamentally aligned with the notion of calibration. Specifically, the problem is cast as one of understanding how well visual judgements based on isolated depth cues, or specific combinations thereof, calibrate with ground truth. While the approach of understanding visuo-motor calibration within a veridical representation model has certainly been critical and influential in helping us understand human capacities in judgements of 3D spatial properties, as well as the efficacy of different cues, it has struggled to capture the complexity of human conscious awareness of space and explain the underlying encodings supporting it.

Other approaches aligned with the prevailing model of 3D vision derive, not from the psychophysical problem of visual calibration (matching visual judgements to spatial ground truth), but from the psychophysical problem of recognition, categorization or shape matching (e.g. Biederman, 1987). The problem of recognition can also be cast as one of probabilistic

inference, where instead of assessing the match between quantitative judgements and ground truth, the problem is one of assessing the match of a categorical judgement with the pre-specified categorical classification of the physical scene (see, for example, Knill & Richards, 1996).

The content of visual representation in the prevailing approach can therefore be viewed as a hierarchy or map of values (to guide motor actions and visual judgements) and symbols representing physical categories (to guide object recognition; Fig 3b)

### **Visual representation, visual encoding and phenomenology of perception**

How do we reconcile this prevailing view of 3D visual representation with the contents of our visual awareness? What we perceive when we look around the visual world is a complex, continuous visual presentation of 3-dimensional objects and surfaces inhabiting the space before us. There is nothing directly quantitative or symbolic about these contents. We do not perceive values or symbols, but rather have awareness of spatial constructs such as distance, space, surfaces, objects, scale, and so on. These contents of conscious visual awareness are a direct communication of the self-directed anticipatory *meaning* of these spatial constructs (Vishwanath, 2005; Albertazzi et al., 2010). Quantity and magnitude are no doubt embedded in this content, since we are able to guide actions and make quantitative judgements on the basis of our awareness, but they are not apparent in the immediate contents of spatial awareness.

The prevailing model's *implicit* view is that the primary contents of visual representation are quantitative values of first- and higher-order geometric attributes (point coordinates, surface normal orientations, etc.) or symbolic attributes (shape identity, category, etc.); see Fig 3B. The content of visual representation under such a view does not directly explain the nature of visual entities the human agent consciously perceives (such as space, surface, object, continuity, scale, etc.) or their agent-centred meaning and intentional significance. Instead, it simply ascribes a quantitative, geometric or categorical label to them. What is therefore required in such a model is an internal higher-order *homunculus* that "reads out" the quantitative values and symbolic identities in order to "paint in" the space, surfaces and objects to our conscious visual awareness and assign meaning and intentionality to them (Fig 3C). A supporter of the prevailing model could, of course, argue that these visual properties and content that are unmodelled in existing treatments are implicitly assumed by the Bayesian framework. However, this begs the question of how we can understand the nature of the encoding of this content within such a model.

An analogy regarding content in the prevailing model can be drawn with aspects of representation in a personal computer. Consider the representation of the text displayed on the computer screen in front of you. A quantitative (numerical) encoding internal to the computer generates coordinates of pixel distributions making up each text character that are appropriately rendered on the screen by the display driver. The internal software encoding might consist of higher-order symbolic constructs (e.g., ascii code, font specifications, etc.) that compactly summarize a collection of adjacent pixels (a character or a word), but there is no sense in which the “meaning” of the letters, words, sentences or paragraphs is embedded anywhere in the computer, either the software or the hardware<sup>2</sup>. The true content is only known to the homunculus at the implementation stage (the software engineer) and the homunculus at the viewing stage (the user).

A more considered approach is to acknowledge that constructs like “distance” and “surface” not only possess agent-centric meaning and content, but, in addition, have a constitutively embedded content of anticipation and agency. We sense the distance to an object not as a quantitative value, but in an anticipatory feeling rooted in motor agency. Our conscious awareness of a 3D surface is not simply the identification of the locus of points at a physical phase-change interface (the definition of physics) but a perceptual object that has an embedded component of agency: the intrinsic anticipation of how such a surface will interact with tactile exploration or manipulation (Vishwanath, 2010; for a more general treatment of embodiment in perception, see Maturana & Varela, 1987).

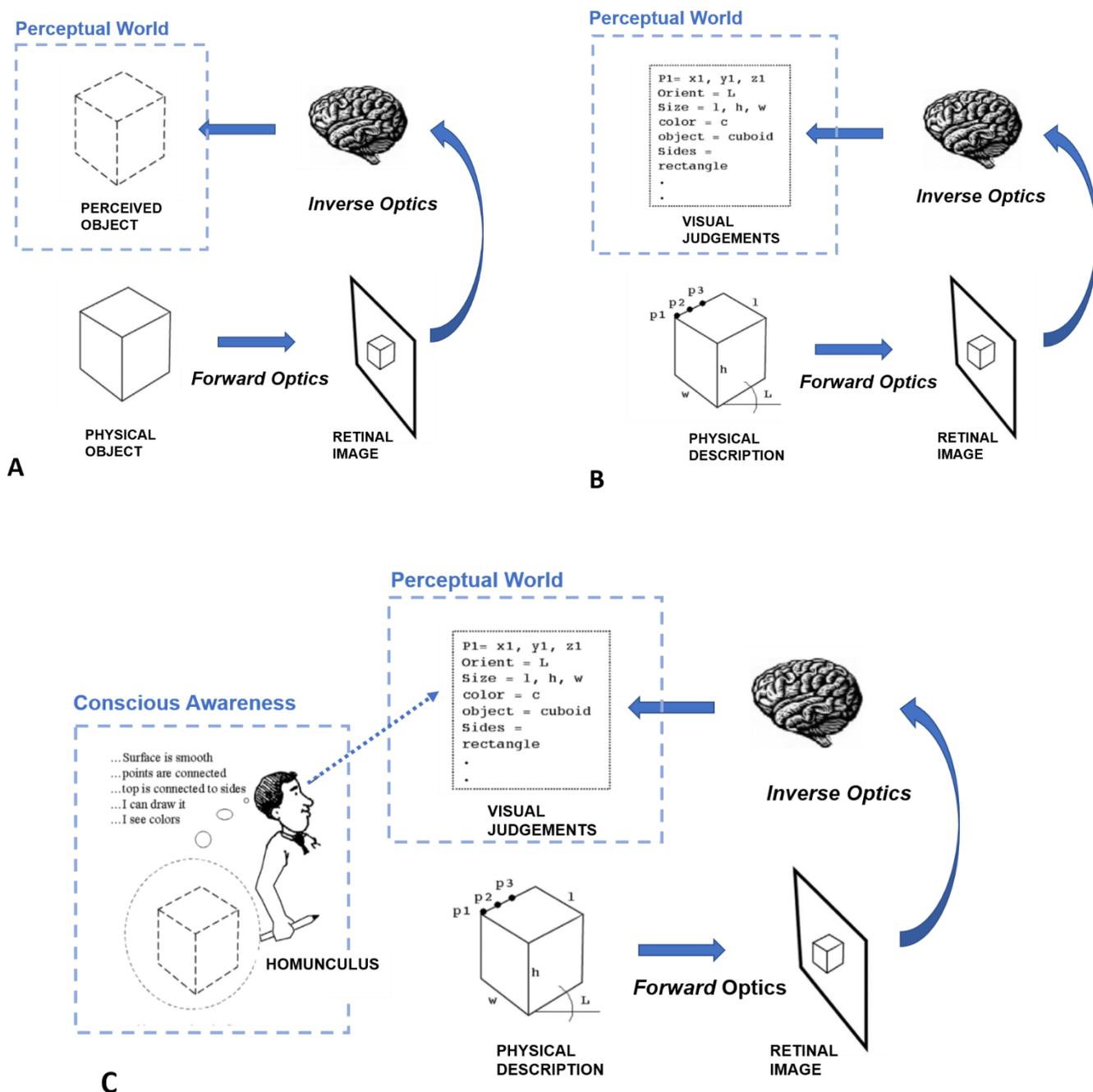
As previously stated, this does not preclude the fact that “quantities” (magnitudes) are implicitly embedded relationally within the structure of this encoding. Instead, the argument here is that a quantitative psychophysical analysis directed at understanding psychomotor calibration or recognition alone (as pursued in the prevailing approach) ends up casting content in terms of the operational variables of psychophysics rather than the content of awareness.

Instead, developing models of the encodings that support such content will require the kinds of *generative* and *nested* structures championed in Leyton’s generative theory of shape (Leyton, 1992). In Leyton’s work, perceptual content goes beyond the constructs of *quantity* and *identity* that are the mainstays of representationalist and inferential models, and considers as a primary constituent the *causal* content of perception. More generally, in moving beyond the prevailing models, it is beneficial to replace the term “representation(s)”

---

<sup>2</sup> This applies equally to word processing programs that have sophisticated capacity to run spell and grammar checks, because the ultimate content and meaning of the internal representation is simply numerical or symbolic with the end goal of generating the pixel coordinates for display.

with the term “encoding(s)” when referring to the structures that give rise to our conscious awareness of a visual space.



**Figure 3** **A.** The standard inverse optics model of 3D perception on which the prevailing models of perception are based. **B.** The prevailing calibration and recognition model of 3D perception, which render perceptual content, in effect, to quantitative values and symbolic labels representing geometric attributes or entity category. **C.** Accommodating the contents of conscious visual awareness in the

prevailing models. Since the quantitative values and symbolic tags carry no specification of the actual content in visual awareness, the model requires a homunculus to interpret and “paint in” the visual scene to consciousness.

## **Alternative Approach**

### *Analysing contents of the visual awareness of 3-dimensionality*

A complete theory of the encodings and information content of conscious visual awareness of space requires not only an understanding of the primitive aspects of 3-dimensionality but also of the constituents of visual space: shapes, surfaces, objects, part/whole relations, etc. An analysis of these aspects (shape, surface, etc.) is outside the scope of the current paper, where the main aim is simply to delineate psychological primitives of 3-dimensionality, particularly, distance, depth and size (scale).

An alternative to the prevailing approach is to take, as the starting point, a detailed analysis of the contents and nature of our conscious awareness of 3-dimensionality. An important first step is to acknowledge that 3D perception is an intentional subjective construct and not a referent to an objective external reality (Vishwanath, 2005, 2010). From a more contemporary perspective, this amounts to the claim that encodings that underlie the perception of 3D space are inherently anticipatory (Albertazzi, 2017; Vishwanath, 2018). They are neither passive representations of (or referents to) an objective physical structure nor an objective communication of physical properties of objects “out there”. Instead, what these encodings provide to conscious awareness is essentially a “plan” about how “what is out there” can be interacted with by the agent thorough motoric or mental operations. The encodings constitute an essentially subjective communication that specifies the conditions and constraints available to behaviour given the sensory motor competencies of the agent.

In a similar vein, individual depth cues must not be viewed simply as contributing to a single objective “master representation” (e.g., Landy et al., 1995). Instead, each cue constitutes a potentially unique manner of encoding of information (content) and each cue likely contributes in idiosyncratic ways to multiple higher order encodings.

This approach allows for a non-teleological view of the evolution of vision, where each evolutionary step is not seen as working toward an increasingly accurate and more resolved objective representation of the scene, but instead, the development of a richer suite of behaviourally relevant visual encodings of the external environment peculiar to the agent. The psychophysical and cue-integration approaches will no doubt continue be critical to validating and guiding the development of models based on analysis of phenomenology.

### *The Experimental Phenomenology Approach*

The alternative approach begins, not with depth cues, but with the first-person phenomenological account of the competencies in the conscious awareness of 3-dimensionality. In other words, we start by enumerating the varieties of the conscious awareness of 3-dimensionality and analyse how they can be systematized. In 3D space, we can distinguish among four main competencies<sup>3</sup>. Later, we will see how these four competencies also have an evolutionary rationale in terms of a phylogenetic development of spatial capacities.

(1) The awareness of objects as being ordered in space in respect to their distance from the observer. The agent simply has an impression of space as partitioned into objects ordered from nearest to farthest with no impression of distances to, or between, objects. Objects themselves are constituted simply by a silhouette, patch or outline with no awareness of the extension of the object (or parts thereof) toward or away from the observer. Human observers have this kind of awareness viewing a simple silhouette image (Fig 4A).

(2) The awareness of 3-dimensional (3D) shape and layout. The agent has an impression of objects laid out in a 3D space, where each object has an overall 3D shape constituted by smaller parts or surfaces of different shapes. Object shapes appear to be more than just silhouettes in that there is an awareness of an extension of the object in depth (away or towards the observer). In quantitative terms, this can be operationalised as knowledge of the *ratios of distances*, or alternatively, the encoding of distances up to a uniform scaling factor (Fig 5, top panel). Human observers have this kind of awareness viewing a pictorial image of a 3D scene (Fig 4B).

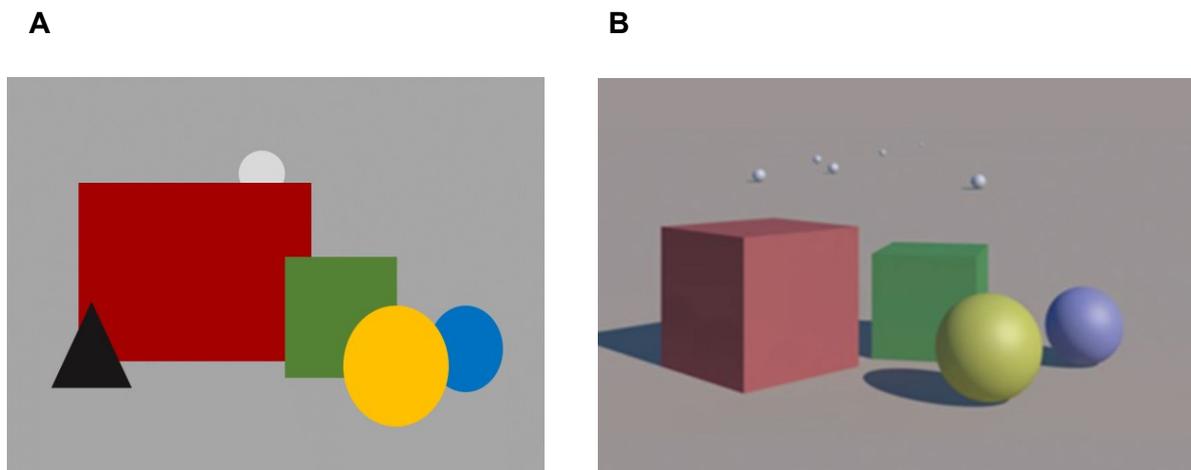
(3) An awareness of the distances of objects from the agent in a form that anticipates adaptive behaviours such as reaching, throwing or walking to a target. This is often referred to as the perception of egocentric distance. In quantitative terms, this can be operationalised as knowledge of absolute or scaled distance between the observer and the

---

<sup>3</sup> The term “competencies” is similar in intent to the original German term deriving from Gestalt Psychology “Aufforderungscharakter” from which J. J. Gibson’s English usage “affordance” derives. However, I have chosen not to use the term affordance here because it is associated specifically with Gibson’s approach whose underlying assumptions (see Vishwanath, 2005) are not compatible with important elements of the conjecture presented here.

object (Figure 5, middle and lower panel). Human observers have this kind of awareness viewing an object in real space.

(4) An awareness of the separation within and between objects in a form that anticipates adaptive behaviours such as grasping. In quantitative terms, this can be operationalised as knowledge of “absolute” or scaled distances between points (Figure 5, middle and lower panel). Human observers typically have this kind of awareness viewing real objects in peri-personal space (<2m) or in stereoscopic images.

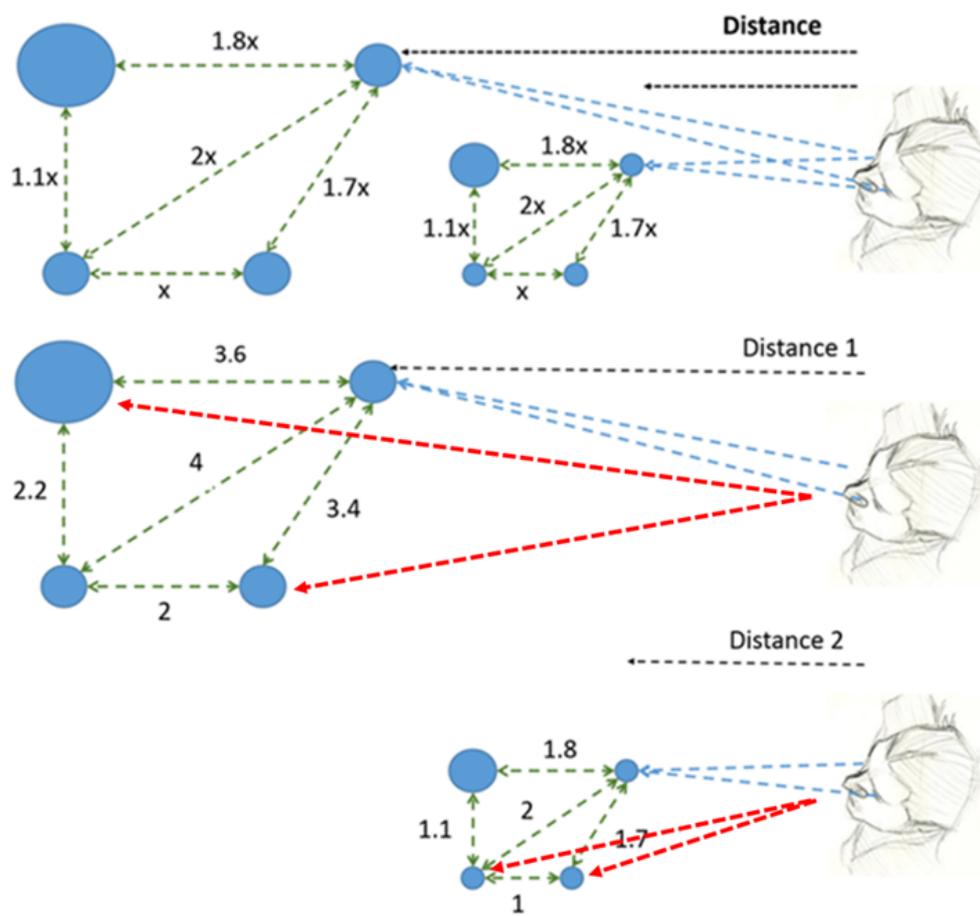


**Figure 4**

A. Awareness of Depth order. B. Awareness of 3D shape and layout

These four competencies in the awareness of visual space do not necessarily appear controversial when considered in the light of the prevailing model of 3D vision which proposes that the primary representation is that of the scaled (egocentric) distances to points in the scene (competency #3; e.g. Landy et al., 1995). This entails knowledge of the scaled distances between points in the scene simply as the difference in egocentric distance to each point, which implies knowledge of the scale of things in the scene (competency #4; Fig 5 middle and lower panel). Furthermore, since distances between points in the scene are known, this entails knowledge of the layout of points up to a uniform scaling factor (ratios of distances between points) implying knowledge of 3D shape/layout (competency #2; Fig 5, top panel). Knowledge of the 3D shape and layout along with observer position in the scene entails knowledge of the depth ordering of objects or shapes (competency #1).

But the argument that all these competencies arise from a single primary representation, appears implausible from an evolutionary standpoint. The implication that the visual system generates one master competency (#4) by reconstructing an objective representation of the scene, and that all other competencies (#1, #2, #3) flow from this, neglects the likelihood that each of these competencies almost certainly evolved idiosyncratically. If we start by assuming that any biological agent that has the capacity to plan action based on visual input has some rudimentary conscious awareness of a “space of operation”, then there are several possible ways in which this evolution might have played out depending on the limitations of the sensory-motor apparatus, niche and evolutionary lineage of the agent.



**Figure 5**

Psychophysical operationalization in Euclidean space of the perception of relative and absolute (scaled) depth and egocentric distance. Top panel: Diagrammatic representation of relative depth perception (perception of 3D object layout and shape). The distance to the fixated object and the scale of the scene are unspecified. The observer perceives a specific layout up to a scaling factor but whose awareness does not distinguish between the two configurations. The observer does not perceive the egocentric distances of the objects or the actual spatial separations between them. Middle and lower panel: Two examples of cases where the observer perceives spatial scale and the

separation between objects. The distance to the fixated object is specified and so is the scale of the configuration. The observer perceives the spatial separations between the objects. Assuming an internally consistent representation, this also implies that the observer is aware of the egocentric distances of objects (red dashed lines).

One plausible rudimentary competency is that of the visual awareness of the distance to an object, either for approach behaviour (such as feeding) or avoidance/retreat behaviour (in the presence of predator). While conscious visual *awareness* of distance<sup>4</sup> is probably not required for rudimentary navigation (e.g., in organisms like the ant), it would be critical for real-time planning of visually guidance for navigation. Another rudimentary competency would be the partitioning of the visual scene into an ordering of visual elements (patches, silhouettes) in depth away from the agent. “Depth ordering” is not to be understood here simply in terms of the mathematical construct of an ordinal variable, but as a complex anticipatory encoding that facilitates planning of behaviour.

These two early competencies may have evolved from distinct subsystems and may have their own idiosyncrasies. For example, it is possible for an agent that has both competencies (awareness of distance to objects<sup>5</sup> and ordering of objects in depth), that these competencies will not always be internally consistent under all contextual conditions. The competency to be aware of the distance to an object need not imply awareness to the distance of all or most objects in the visual field. It might only exist for the object of interest. The agents might have conscious awareness of the distance to a single object (at the point of gaze or attention) and simultaneously only the depth ordering of the other surrounding objects, where scanning to a different object would bring its “distance” into awareness. Such “non-objective” states may not be normally evident to the agent in its overall spatial awareness due to seamless shifts in localised awareness of object distance when scanning from object to object; in a similar way in which human observers are typically unaware of the idiosyncrasies in awareness of space outlined in the introduction.

An agent higher in the evolutionary tree, in addition to the above competencies, may also have awareness of the relative layout and 3D shape of objects in space which might facilitate adaptive behaviour such as spatial planning along with recognition and categorization of 3D object shape. It is plausible that this competency is an offshoot of the

---

<sup>4</sup> The term “distance” here is not intended simply as an objective quantitative estimate, but a subjective anticipatory perceptual attribute associated with the motor competencies of the agent. See below how distance should be considered an embodied attribute.

<sup>5</sup> The term object is used loosely here. It does necessarily imply the capacity to be aware of 2D shape, but simply a segmented portion of the visual field.

encodings that underlie awareness of depth ordering (competency #1) and remain distinct from the competency of awareness of distance (#3).

Finally, certain agents who have motoric capacities for fine-grained manual manipulation or advanced predatory planning may, in addition, have developed encodings that allow for awareness of the extensions of objects, or the space between objects, in a format that anticipates behaviours such as grasping (competency #4; scale and absolute distances). Once again, this competency will plausibly have evolved to be distinct from the other three, even though it may receive inputs or have outputs to the other encodings.

In summary, it is highly probably that the 4 basic competencies in spatial awareness (distances to objects, distances among objects, spatial layout / object shape and the ordering of objects in depth) most likely evolved as distinct encodings. If we follow this logic, then what the human agent is consciously aware of in terms of its “space of operation” must also have dissociated representations and idiosyncratic features, notwithstanding the fact that, under normal operating conditions, the (human) agent’s awareness is that of a unified and objective space (the feeling that one is peering out into an observer-independent physical space).

The three phenomenological observations described at the very start of this paper demonstrate that scrutiny of our visual awareness of space under different stimulus conditions reveals cracks precisely at the boundaries where we would expect from a principled classification of subjective competencies in visual awareness. For example, pictorial perception appears to reveal a dissociation between the awareness of object layout and 3D shape (competency #2) and the awareness of the tangible extensions of objects and spatial separation (the feeling we can reach out touch or pick things up) relating to competency #3. A similar dissociation in awareness is revealed in difference in monocular and binocular viewing of real scenes, as well as in changes in spatial awareness with distance. These observations led to the proposition that awareness of visual space is underwritten by two distinct spatial encodings (Vishwanath, 2010, 2014)

### **Dual-encoding model of 3D perception**

The conjecture of a dissociation in conscious visual awareness of 3D space arose from the consideration of the differences in phenomenology in the perception of pictorial and real (or stereoscopic) scenes (Vishwanath, 2010, 2014). The conjecture proposed that the awareness of spatial layout and 3D shape is underwritten by an exocentric encoding containing information operationally related to depth ratios and so expected to lack an

awareness of spatial scale (Fig 5a). The awareness of scale and spatial separation between and within objects was conjectured to be underwritten by an encoding containing information operationally related to egocentric distance and egocentrically scaled depth (absolute depth). This encoding is hypothesised to underlie the phenomenological sensation of the solidity and tangibility of objects and the impression of a “real” spatial separation or negative space between things (the impression referred to by the term *stereopsis*). These phenomenological attributes give rise to an overall sensation of reality, spatial immersion and presence; consistent with Michotte’s (1949) claim that the phenomenology of stereopsis is associated with capacity for visually guided manipulation. This dual-encoding conjecture provides an explanation for the ubiquitous duality of pictorial depth perception (Vishwanath & Hibbard, 2013).

Furthermore, the conjecture proposed that the subjective impression of object tangibility and real spatial separation between objects lies on a continuum from conditions where it is entirely absent (pictorial images viewed from a near distance with both eyes) to conditions where it appears strongest (real or stereoscopic objects viewed with both eyes in personal and peri-personal space). This continuum is proposed to be correlated with the quality (reliability) of visual information specifying egocentric distance in real scenes. Objects at close distances are expected to produce a strong phenomenological impression of tangibility and spatial separation while at farther distances (where distance information is less reliable) the conjecture predicts an exponential reduction, such that objects in far vista space appears almost pictorial (Vishwanath, 2010,2014). A range of psychophysical and qualitative empirical evidence has been presented in support of this conjecture (Vishwanath & Hibbard, 2013; Vishwanath, 2014; Vishwanath 2016; Volcic et al., 2013).

The original proposal attributed changes in the phenomenological strength of the impression of spatial separation (*stereopsis*) to statistical reliability or certainty in estimates of scaled depth. However, cast in the framework of the Intrinsic Constraint (IC) Model of depth cue integration (Domini et al., 2006) changes in phenomenological impression of spatial separation (reduction with distance) can be directly attributed to gain (magnitude) of the depth estimate rather than its reliability (see Domini & Vishwanath, 2020, for a review). Under this interpretation, it is not the certainty of depth estimates that degrades with distance, but the estimates of scaled spatial separation (scaled depth) become exponentially smaller for greater distances beyond personal space (regardless of the actual physical separation) due to the reduction in informativeness of the sensory information specifying distance. This interpretation based on the IC model is more consistent with some important aspects on spatial phenomenology.

Based on the known division of the visual pathways into the ventral (temporal) and dorsal (parietal) streams of processing, the dual-encoding conjecture implies that awareness of spatial scale and egocentric distance are most likely instantiated in encodings in the dorsal stream, since they are involved in direct visuo-motor guidance, while those underlying awareness of (unscaled) spatial layout and object shape that only indirectly guide action are instantiated in the ventral stream (Vishwanath, 2010). Thus, the substrate underlying visual phenomenology associated with stereopsis (awareness of spatial separation, object solidity, tangibility, realness and immersion/presence) is predicted to be in the dorsal visual stream. This prediction also aligns with the suggestion in Hibbard (2008) that the phenomenology of stereopsis may be associated with the dorsal rather than ventral visual stream despite a prevailing view that the former is not involved in conscious vision.

Recent neuroimaging evidence (Uji et al, 2018, Uji et al., 2019) is consistent with this prediction of the dual encoding model. This research has shown that contrasts between stimulus conditions where there is the impression of tangibility and real spatial separation (stereoscopic viewing, monocular aperture viewing of pictorial images) and conditions where this impression is absent despite the perception of 3D shape (normal binocular viewing of pictorial images) reveals selective activation of dorsal visual areas, specifically in the posterior parietal cortex (Uji et al, 2018, Uji et al., 2019).

### **Tripartite encoding models of spatial awareness**

An implicit assumption of the dual-encoding model of 3D space perception is that awareness of egocentric distances to objects and scaled distances (separation) between objects both arise from the same encoding, since they are operationally interchangeable under the assumption of single internally consistent representation (Figure 5B & C). However, both the psychophysics and the phenomenology of space perception beyond personal space into interaction and vista space (described in the introduction) complicates this picture.

Phenomenological and psychophysical observations point to a clear dissociation between awareness of egocentric distance and relative depth in the action space of the observer (2m-20m; Loomis et al., 1995). This appears consistent with the dual encoding model which conjectures separate encodings of relative depth (3D shape and layout) on one hand and egocentric distance and scaled depth separation on the other (Vishwanath, 2014). The dual-encoding model proposes that the awareness of real spatial separation (stereopsis) is associated with the latter encoding, and is what underlies the phenomenology usually associated with stereopsis (object solidity, tangibility, impression of negative space)

However, empirical evidence suggest that the dual encoding model cannot fully capture the nature of spatial awareness in its entirety.

### *Empirical evidence supporting a tripartite model*

Empirical evidence suggests that egocentric distance can be accurately estimated to at least 20m and there is surprisingly little drop in the variability (standard deviation) of blind walking responses up to distances of at least 12m (Loomis et al., 1992). If the visual system has accurate and reliable distance information up to 20m and reliability of disparity is independent of viewing distance (since it is a purely retinal cue), we should predict that scaled depth estimates should not appreciable reduce in reliability (or gain) for distance between 2 to 20 meters. According to the conjecture in Vishwanath (2014), this should imply no significant reduction in the phenomenological impression of spatial separation at least up to 20m. However, judgments of the subjective impression of real separation in depth reveal a rapid decline with viewing distance, such that the subjective impression of separations between objects several meters apart viewed from a distance of even 15m is significantly attenuated compared to objects separated by just 30cm viewed from a meter away (Vishwanath, 2014). If we accept that the subjective awareness of spatial separation is linked to derivation of scaled depth, then this suggests that encoding of egocentric distance and scaled depth are paradoxically dissociated; contrary to what is indicated in fig 5B, C.

The dissociation between scaled depth separation and egocentric distance is further highlighted by other empirical data. First, blind walking responses under monocular viewing appear not be any less accurate than under binocular viewing (Ooi & He, 2015) despite the fact that monocular viewing of objects at distances of several meters shows a dramatic loss in the impression of spatial separation compared to binocular viewing. Also, blind-walking responses in strabismic observers who lack functional binocular vision is largely comparable to individuals with typically developed binocular vision (Ooi et al., 2015), even though their impression of stereopsis (spatial separation) is weak or absent (Barry, 2009).

A further complication is that the capacity to make motor actions that rely on egocentric distances appears to depend on the distance at which such actions are made. Studies with strabismics observers on tasks in reach space that require egocentric distance information show deficits (O'Connor et al., 2010a, 2010b) compared to typical stereovision observers despite accuracy in blind walking tasks in action space (>2m; Ooi et al., 2015). Similarly,

typical stereovision observers show deficits in tasks for monocular compared to binocular viewing in reach space (Henson & Williams, 1980; Tidbury et al., 2014), even though, as previously mentioned, blind walking responses to distances of several meters show no significant differences (Ooi et al., 2015).

These observations point to an important dissociation between egocentric distance coding in near space (which appears to significantly rely on binocular vision) and distance coding in action space (which appears to not be reliant on binocular vision). Moreover, conditions which yield accurate motor responses in near space is associated with subjective awareness of real spatial separation, but awareness of spatial separation (stereopsis) in action space is diminished despite the capacity for accurate locomotion revealing a dissociation between encodings that underlie awareness of spatial separation and those underlying awareness of egocentric distance.

### *Tripartite encoding and phenomenology*

Considered in the context of the proposal in Vishwanath (2014), this suggests that spatial awareness is underwritten by a tripartite dissociation among encodings specifying relative depth, scaled depth and egocentric distance. The phenomenological and empirical evidence suggest the following tripartite encoding model:

- I. Encoding of relative spatial layout and 3D shape (relative depth) that is distance independent
- II. Encoding of egocentric distance and scaled depth (scaled within object-extension and between object separation) optimised for near viewing only (<2m)
- III. Encoding of egocentric distance optimised for ambulatory distances (action space >2m) but without encoding of scaled depth.

Note that the proposal does not strictly restrict encodings II and III to specific regions of the visual field, but simply conjectures that they are optimized for specific regions. Thus, while scaled depth encoding is conjectured to be optimized for near space, it is likely to extend beyond this but with an accelerating degradation in terms of accuracy and/or precision with increasing distance. Vishwanath (2014) conjectured that the characteristic phenomenology of stereopsis (the impression of object solidity, tangibility, realness and the impression of real spatial separation (negative space) was due to the encoding of scaled depth and distance (encoding II). Since this encoding is optimized for near space, the conjecture can explain why the phenomenology of stereopsis and particularly the awareness of real spatial separation is strongest in the personal space of the observer, but also present (in

substantially diminished degree) at distances up to or beyond 100 meters (Vishwanath, 2014). Here we further conjecture that the phenomenology of immersion and presence—the feeling that one is located in an interactable real space—is generated due either or both encodings II and III. This can explain why the feeling of immersion and presence does not require signals (binocular disparity, motion parallax) that underlie derivation of scaled depth (encoding II) but is present even when the observer is static and viewing a real scene monocularly. In this case derivation of egocentric distances at the ambulatory scale (encoding III) is still possible (Loomis et al. Ooi & He, 2015). This is evident in the fact that closing an eye does not noticeably alter the feeling of presence or immersion, unless one pays careful attention to the attenuation of spatial separation in near space and beyond due to the loss of binocular disparity. Finally, a pictorial scene viewed with both eyes does not typically generate an impression of either stereopsis or immersion/presence because there is only an awareness of relative depth (encoding I). However, special viewing conditions (monocular-aperture viewing, synoptic viewing, viewing images with high resolution screens at a distance, etc.) can induce a noticeable impression stereopsis (Vishwanath & Hibbard, 2013; Vishwanath, 2013). Other stimulus condition (e.g., full-scale dioramas where foreground elements are real but background and distant vista is painted on a curving backdrop) can induce a substantive feeling of immersion/tangibility.

#### *Tripartite encoding, adaptive significance and neurophysiology*

The logic of a tripartite model is supported by a consideration of the differences in adaptive significance of each competency. Awareness of object shape and relative layout is useful for planning tasks requiring identification, recognition and visual orientation. Awareness of scaled within-object extension and inter-object separation is critical for planning motor behaviour such as object manipulation and organization but not necessary for discerning the shape of objects or relative layout. Moreover, such encodings would only be adaptively significant within personal (reach) space within which such behaviours occur. They are less critical in action space and beyond where encodings that support planning of more coarse-grained ballistic actions (e.g., throwing) and locomotor/ambulatory actions such as visually guided approach, retreat and navigation will be more useful. Such planning requires only an encoding of object distances.

The logic of the tripartite dissociation, particularly a dissociation between encodings supporting judgements of scaled depth in near space and egocentric distance in far space, is also revealed in the sensory signals available considered within a psychophysical operationalization. Fine-grained absolute or scaled depth judgements are thought to rely on

combining information from retinal depth cues (such as disparity, motion parallax) and egocentric distance cues (primarily, ocular vergence). But ocular vergence is thought to be largely non-informative beyond personal space (>2m; Tresilian et al., 1999)<sup>6</sup>. On the other hand, tasks requiring judgements of egocentric distance to targets beyond 2m (e.g. blind walking to previewed targets) are thought to rely on ground plane information and declination from eye level (Loomis et al., 2002; Ooi et al., 2001) which are useful for distances greater than 2m and support relatively accurate judgments for ambulatory distances out to at least 20 m (Loomis et al., 1992).

In addition to the above functional arguments for a dissociation, considerations from the evolution of sensory motor systems also support the likelihood of dissociated encodings underlying egocentric distance and scaled depth perception. As previously mentioned, it is likely that the earliest competency in awareness of visual space would be that of egocentric distance, which is a crucial component of visually guided navigation. Many aspects of navigation in organisms with less complex nervous systems (e.g., insects) could rely primarily on memory mechanisms underwriting functions such as dead reckoning or path integration (Wehrner et al., 1996). However, online planning of navigation based on real-time visual input would, at minimum, require awareness of distances to at least the object of interest (even if the visual encoding of the objects themselves were quite rudimentary and did not specify the 3D shape of the object). More advanced behavioural planning would additionally benefit from an awareness of spatial layout (depth ratios). Finally, only the most behaviourally advanced organisms which require fine-grained planning of manual actions in personal and peri-personal space would benefit from spatial awareness of scaled depth.

In terms of neurophysiology, the well-established distinction in mammalian brain between the temporal and parietal divisions of visual cortex (Goodale & Milner, 1992; Mishkin, Ungerleider & Macko, 1983) provide a logic for the dissociation between encodings supporting awareness of object shape and layout and those supporting fine-grained manual motor actions. A longstanding view of this dissociation was that one set of encodings (temporal/ventral) was responsible for conscious perception while the other (parietal/dorsal) was only used to subconsciously guide action. However, more recent accumulation of evidence suggests that the dorsal stream also plays an important role in conscious perception (e.g., Freud et al., 2014). This is consistent with the proposal here, in Michotte (1949) and in Vishwanath (2010) that the consciousness perceptual impression of reality is

---

<sup>6</sup> Recent evidence showing no modulation of distance responses with changes in vergence in even personal space (<1m), has questioned whether there is any role for vergence in distance perception (Linton, 2020)

likely associated with scaled encodings required for guiding manual actions which implicate the dorsal visual stream.

The role of the parietal cortex in encodings that underlie visually guided manual action and the role of temporal cortex in object shape coding is well established. Potential substrates for the visual encoding of distances at ambulatory scales are less well understood.

However, evidence supports the view proposed here, namely, that visual awareness of egocentric distance at the ambulatory scale is distinct from encodings supporting scaled representations required for manual action planning in near space.

Much of the evidence of spatial planning at the ambulatory scale has implicated the medial aspects of the inferior temporal cortex (entorhinal cortex and parahippocampus) in functions related to memory-guided path integration in rodents and spatial (but not object) perception in humans. However, evidence from rodent studies (and limited human studies) also implicates the posterior parietal cortex (PPC) in navigation (e.g., Commins et al., 1999; McNaughton et al., 1989; Nitz, 2006; Rosenbaum et al., 2004); not on the basis of direct connections from visual cortex, but indirectly from efferent inputs from medial temporal areas known to be involved in navigation (Whitlock et al., 2008). The PPC is known to have numerous inputs from entorhinal cortex and is thought to be involved in converting spatial coding instantiated in the entorhinal cortex into action relevant encodings (Whitlock et al., 2008). One prevailing view is that relevant PPC neurons “do not express perceived space...instead, the observations point to a possible role for PPC neurons in planning and execution of navigational behaviours” (Whitlock et al., 2008).

However, the conjecture made here is that visual awareness of distance does not arise simply from a location map but is an anticipatory and embodied construct directly underlying spatial planning, implicating the network linking medial temporal and parietal areas in the visual awareness of distance rather than just subconscious control of motor actions.

Particularly intriguing in this regard is the still debated function of grid cells first discovered in rat entorhinal cortex. Two aspects of the neurological evidence provide support for the view that these medial temporal areas could contribute to encodings that support the conscious awareness of egocentric distance at the locomotor scale, particularly the embodied and anticipatory aspects of this awareness.

First, grid cells are part of paleocortex structures in temporal cortex (entorhinal cortex) that efferent to the more primitive divisions of the cortex (archicortex) associated with encoding spatial memory (hippocampus). In humans, parahippocampal areas of the temporal cortex have been associated with encoding of spatial layout in fMRI studies, with highest activations for images of outdoor scenes, but not 3D objects alone or relative layout of

objects without spatial context (Epstein & Kahnwisher, 1998). Importantly, they do not appear to be implicated in recognition or memory (Epstein et al., 1999).

Second, grid cell organization has been shown to be altered by the shape of the testing enclosure, challenging the original interpretation of grid cells as providing a uniform and fixed metric for memory-guided spatial navigation (Krupcic et al., 2016). One possible interpretation of this data is that because grid cell layout is dynamically sensitive to visual input underlying encodings of distances, grid cells may be part of a system that underlies the anticipatory awareness of distances to locations in space; rather than being only involved in memory-based path integration or route finding; even though they may be the input to such memory based integration mechanisms in hippocampus.

Taken together, studies in rodents and humans are consistent with the idea that areas in parahippocampus and entorhinal cortex are engaged with rudimentary constituents of visual space of the kind required to develop anticipatory encodings relevant to an ambulatory scale, rather than 3D object shape. If so, whether these areas alone, or in conjunction with PPC, are responsible for the visual awareness of distance remains an open question, especially because much of the literature implicates entorhinal areas in allocentric maps rather than egocentric encoding. However, studies that have examined allocentric vs. egocentric encoding in these regions (particularly in rodent studies) have typically used the term 'egocentric' in terms of perceived direction rather than distance. Furthermore, they have been operationalised for memory related tasks rather than perceptual tasks. The role of networks in this area from a perceptual standpoint has been much less studied.

It is plausible that the more primitive aspects of the cortex like the parahippocampal areas and entorhinal cortex which are implicated in spatial competencies required for navigational planning were arguably among the earliest to evolve in terms of spatial awareness given their location in archicortex; and as noted earlier, it is logical to believe that the earliest competency of spatial awareness is that of object distance. An important point of the conjecture in this paper is that distance perception is not simply knowledge of a quantitative value (the egocentric coordinate of the object location) but, rather, is an *anticipatory* awareness of a location embedded within a complex sensory-motor encoding of space. This alternative view of egocentric distance perception is what makes structures like the grid cell, which appear to dynamically incorporate both sensory and locomotor components, potential substrates for what underlies real-time conscious anticipatory awareness of distance.

This view is particularly persuasive in considering the classic blind-walking paradigm. The subjective phenomenology of the task suggests that the observer has a conscious anticipatory visual encoding of distances, which, in the blind-walking response, involves a

sort of “embodied cancellation” of the anticipated distance based on the internal record of distance traversed, as the observer ambulates with their eye’s closed. (The reader is encouraged to experience the blind walking task for themselves in a safe location). It is clear upon doing the task that it is not as though the observer cognitively notes a quantitative estimate from visual perception, and then applies this to derive the number of steps or duration required to blind-walk a matched distance. Instead, the task has an inherent anticipatory and embodied aspect to it. The nature of grid cell firing logic, in that they “mark” traversed distances, but also that the spacing of the marked positions (nodes) appear to be dependent on sensory input (Krupcic et al., 2016) is suggestive of a mechanism that could contribute to this competency.

## **Summary and Conclusion**

Integrating phenomenological observations, empirical data, evolutionary logic and neurophysiological evidence leads to the conjecture that human conscious awareness of visual space is underwritten by three major distinct spatial encodings. The most primitive of these is proposed to support the competency of the conscious awareness of distance (operationally, egocentric distance) and is hypothesised to be primarily instantiated in temporal archicortex regions. The second encoding supports the competency of awareness of object layout and shape without scale (operationally, relative depth) likely instantiated in ventral visual stream of the neocortex. The third supports the fine-grained competency of awareness of intra- and inter-object spatial separation (operationally, scaled (absolute) depth) and instantiated in the dorsal visual stream. This third competency is what is conjectured to underlie the phenomenology of object solidity, spatial separation, tangibility and object realness. The combined effect of the first (distance) and third (spatial separation) competency is what is conjectured to underly the feeling of spatial immersion and presence.

Furthermore, the intended behavioural “range of operation” appears to differ among these distinct encodings. The scale-free encoding of 3D shape and layout are likely independent of viewing distance. Awareness of fine-grained scaled relations (absolute depth) are likely restricted to personal and peri-personal space, while becoming exponentially ineffective in action space and beyond; underscored by the dramatic drop in the visual impression of spatial separation (stereopsis) between objects in action space and beyond (>5m).

Encodings that underlie our awareness of object distance likely extend from personal space

into the limits of action space, being effective up to at least 20m with a graceful drop off beyond.

In conclusion, what is argued here is that although our conscious visual awareness of objects and space appears to the uncritical eye to be a seamless, unified and objective instantiation of external world, it is most likely constituted by distinct and dissociated encodings that have evolved in a way specialised to the evolutionary development of visuo-motor and mental competencies. They have likely evolved different ranges of operation with no guarantee of consistency among them. Finally, the nature of our conscious awareness of these spaces reveals that these competencies are not simply the recovery of the attributes of physical world but are inherently subjective constructs that provide an intentional and anticipatory interface with which the organism can interact with its niche.

## References

Albertazzi, G. van Tonder, & D. Vishwanath (2010). *Perception beyond inference: The information content of visual processes*. Cambridge, MA: MIT Press.

Albertazzi L. (2017) Microgenesis of Anticipation: Windowing the Present. In: Poli R. (eds) *Handbook of Anticipation*. Springer, Cham. p285-302

Aguirre, G.K., Detre, J.A., Alsop, D.C., and D'Esposito, M. (1996). The parahippocampus subserves topographical learning in man. *Cerebral Cortex*, 6, 823–829.

Biederman, I. (1987) Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94(2):115-147.

Commins, S., Gemmell, C., Anderson, M., Gigg, J., O'Mara, S.M. (1999) Disorientation combined with bilateral parietal cortex lesions causes path integration deficits in the water maze. *Behavioural Brain Research*, 104:197–200

Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout: The integration, relative dominance, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Handbook of perception and cognition: Perception of space and motion* (Vol. 5, pp. 69–117). New York, NY: Academic Press.

Domini, F., & Braunstein, M. L. (1998). Recovery of 3-D structure from motion is neither Euclidean nor affine. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4), 1273.

Domini, F., Caudek, C., & Richman, S. (1998). Distortions of depth-order relations and parallelism in structure from motion. *Perception & Psychophysics*, 60(7), 1164-1174.

Domini, F., Caudek, C., & Tassinari, H. (2006). Stereo and motion information are not independently processed by the visual system. *Vision Research*, 46, 1707–1723.

Epstein, R. & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392, 598-601.

Epstein, R., Harris, A., Stanley, D. & Kanwisher, N. (1999). The parahippocampal place area: Recognition, navigation, or encoding? *Neuron*, 23, 115-125.

Erkelens, C. J. (2015). The extent of visual space inferred from perspective angles. *i-Perception*, 6, 5–14.

Freud E, Plaut, DC & Behrmann, M (2016) 'What' is happening in the dorsal visual pathway *Trends in Cognitive Sciences* 20 (10), 773-784

Goodale M.A. & Milner, A.D. (1992). "Separate visual pathways for perception and action". *Trends in Neuroscience*, 15 (1): 20–5.

Hibbard, P. B. (2008). Can appearance be so deceptive? Representationalism and binocular vision. *Spatial Vision*, 21, 549–559.

Knill, D. C. (2007). Learning Bayesian priors for depth perception. *Journal of vision*, 7(8), 13-13.).

Knill, D. & Richards, W. (1996) *Perception as Bayesian Inference*. Cambridge: Cambridge University Press

Krupic J, Bauza M, Burton S, Barry C, O'Keefe J. 2015. Grid cell symmetry is shaped by environmental geometry. *Nature*, 518:232–235.

Landy M.S., Maloney L.T., Johnston E.B., Young M.J. (1995) Measurement and modelling of depth Cue combination: in defence of weak fusion. *Vision Research*, 35:389–412

Landy, M. S., Banks, M. S., & Knill, D. C. (2011) Ideal-observer models of cue integration. In: J. Tommershauser, K. Kording & M. Landy (Eds) *Sensory cue integration*. Oxford University Press: Oxford, UK. pp 5–29

Lappin, J. S., & Craft, W. D. (2000). Foundations of spatial vision: from retinal images to perceived shapes. *Psychological Review*, 107(1), 6.

Leyton, M (1992) *Symmetry, Causality, Mind*. Cambridge, Massachusetts: MIT Press

Linton P. (2020) Does vision extract absolute distance from vergence?. *Attention Perception Psychophysics*. 82(6):3176-3195.

Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 906–921.

Loomis, J. M., & Knapp, J. M. (2003). Visual perception of egocentric distance in real and virtual environments. In L. J. Hettinger & M. W. Haas (Eds.), *Virtual and adaptive environments* (pp. 21–46). Mahwah, NJ: Erlbaum.

Loomis, J. M., Philbeck, J. W., & Zahorik, P. (2002). Dissociation of location and shape in visual space. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 1202–1212.

Maturana, H. R. & Varela, F.J. (1992) *The Tree of Knowledge: The Biological Roots of Human Understanding*. Shambhala

McNaughton, B.L., Leonard, B., Chen, C. (1989) Cortical-hippocampal interactions and cognitive mapping: A hypothesis based on reintegration of the parietal and inferotemporal pathways for visual processing. *Psychobiology*, 17:230–235

Michotte, A. (1948). The psychological enigma of perspective in outline pictures. In G. Thines, A. Costall, & G. Butterworth (Eds.), *Michotte's experimental phenomenology of perception*. Hillsdale, NJ: Erlbaum. (pp. 174–187)

Mishkin, M., Ungerleider, L. G. & Macko, K. A. (1983) Object vision and spatial vision: two cortical pathways. *Trends in Neuroscience*, 6, 414-417

Niederée, R., & Heyer, D. (2003). The dual nature of picture perception: A challenge to current general accounts of visual perception. In H. Hecht, R. Schwartz, & M. Atherton (Eds.), *Looking into pictures: An interdisciplinary approach to pictorial space* (pp. 77–98). Cambridge, MA: MIT Press.

Nitz D.A. (2006) Tracking route progression in the posterior parietal cortex. *Neuron* 49:747–756.

O'Connor AR, Birch EE, Anderson S, Draper H. (2010a) Relationship between Binocular Vision, Visual Acuity, and Fine Motor skills. *Optometry and Vision Science*; 87(12):942-7.

O'Connor AR, Birch EE, Anderson S, Draper H, Group FR. (2010b) The functional significance of stereopsis. *Investigative Ophthalmology and Vision Science*;51(4):2019-23

Ooi T.L., & He Z.J. (2015) Space perception of strabismic observers in the real world environment. *Investigative Ophthalmology & Visual Science*. 56:1761-8.

Palmer, S. E. (1999) *Vision Science: Photons to Phenomenology*. Cambridge, MA: MIT Press.

Rosenbaum, R.S., Ziegler, M., Winocur, G., Grady C.L., Moscovitch, M. (2004) "I have often walked down this street before": fMRI studies on the hippocampus and other structures during mental navigation of an old environment. *Hippocampus*, 14:826–835

Sauers, J. A. & Knill, D.C. (2001) Perception of 3D surface orientation from skew symmetry. *Vision Research*, 41(24), 3163-3183;

Sekuler, R. & Blake, R. (2006) *Perception*. McGraw Hill.

Svarverud, E., Gilson, S.J. and Glennerster, A., (2012) A demonstration of 'broken' visual space. *PLoS ONE*, 7(3), e33782.

Todd, J. T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences*, 8(3), 115-121.

Tresilian JR, Mon-Williams M, Kelly BM. (1999) Increasing confidence in vergence as a cue to distance. *Proc Biol Sci*. 266(1414):39-44.

Uji, M., Jentzsch, I., Redburn, J., & Vishwanath, D. (2019). Dissociating neural activity associated with the subjective phenomenology of monocular stereopsis: An EEG study. *Neuropsychologia*, 129, 357–371.

Uji, M., Lingnau, A., Cavin, I., & Vishwanath, D. (2019). Identifying cortical substrates underlying the phenomenology of stereopsis and realness: A pilot fMRI study. *Frontiers in Neuroscience*, 13, 646.

Vishwanath, D. (2005). The epistemological status of vision and its implications for design. *Axiomathes*, 15, 399–486.

Vishwanath, D. (2010). Information in surface and depth perception: Reconciling pictures and reality. In L. Albertazzi, G. van Tonder, & D. Vishwanath (Eds.), *Perception beyond inference: The information content of visual processes* (pp. 201–240). Cambridge, MA: MIT Press.

Vishwanath, D. (2014). Toward a new theory of stereopsis. *Psychological Review*, 121, 151–178.

Vishwanath, D. (2018) The anticipatory structure of perception. In R. Poli (Ed) *Handbook of Anticipation*. Springer: Cham.

Vishwanath, D., & Hibbard, P. (2013). Seeing in 3D with just one eye: Stereopsis in the absence of binocular disparities. *Psychological Science*, 24, 1673–1685.

Volcic, R., Vishwanath, D., & Domini, F. (2014). Reaching into pictorial spaces. In B. E. Rogowitz, T. N. Pappas, & H. de Ridder (Eds.), *Human Vision and Electronic Imaging XIX: Proceedings of SPIE*.

Whitlock, J.R., Sutherland, R.J. Witter, M.P., Moser, M.B., Moser, E.I. Navigating from hippocampus to parietal cortex. *Proceedings of the National Academy of Sciences*, 105, 14755-14762.

Wehner, R., Michel, B, & Antonsen, P (1996) Visual navigation in insects: coupling of egocentric and geocentric information. *Journal of Experimental Biology*, 199: 129-140.