Review of Philosophy and Psychology
https://doi.org/10.1007/s13164-025-00780-1

The Hierarchy of Selves in Perception

Catherine Hochman[1]

Accepted: 19 August 2025

Abstract
A prevalent philosophical view holds that information about the self remains implicit in perceptual representation. While it correctly describes the most basic perceptual representations, I argue that this view does not hold for more complex, but still low-level, perceptual representations of space. I propose that in the course of perceptual processing, implicit information about the self is made explicit. My discussion centers on a detailed study of egocentric reference frames. I suggest that within the perceptual system of a creature, these reference frames form a hierarchy, in which each frame that is implicit at one stage of perceptual processing is explicit at the next. This hierarchy allows implicit information about the self to be articulated as processing advances. I show how standard explanations of two perceptual phenomena – gaze shifts and the illusory experience of self-motion – are committed to and thus support this hierarchical picture.

Keywords Self-representation · Perceptual representation · Egocentric reference frames

1 Introduction

Self-representation seems to lie at the heart of self-knowledge, self-consciousness, and agency. As a result, significant work has gone into analyzing *de se* thoughts – thoughts we have about ourselves (Evans 1982; Kaplan 1989; Lewis 1979; Perry 1979). But to truly grasp the nature of self-representation, we should examine its possible origins in more primitive kinds of mental states.

✉ Catherine Hochman
chochman@ucla.edu

1 Department of Philosophy, University of California, Los Angeles, 394 Dodd Hall, 390 Portola Plaza, Los Angeles, CA 90095, USA

Published online: 03 October 2025

Perception provides fertile ground to explore the origins of self-representation. The tight connection between perception and one's self, between what is perceived and who is perceiving, prompts the question of whether and how the self is accounted for in perceptual representation. A prevalent philosophical view holds that the self remains implicit in perceptual representation (Campbell 1994; Ismael 2012; Musholt 2015; Recanati 2007, 2012). The basic idea is intuitive. What I perceptually represent are the objects of perception – things such as colors, textures, and shapes. Because I am not usually among the objects I perceive, I do not perceptually represent myself. Nonetheless, I locate the objects I perceive relative to myself, and so it seems that the self plays some implicit role in perceptual representation.[1]

I'll interpret this view to imply that the self is implicit in all perceptual representations.[2] To better understand this idea, it will help to define the notions of 'implicit' and 'explicit' at play. Let's assume that perceptual representations are structured, meaning that complex perceptual representations are composed of constituent perceptual representations bound together in a certain way.[3] Information is explicit in a complex perceptual representation (or any complex representation) in virtue of being represented by a constituent of that representation. So, information about the color red, for example, is explicit in a complex visual representation of a red apple in virtue of a constituent representation of the color red. By contrast, implicit perceptual information can be understood as information that is not represented by one of the constituents of a complex perceptual representation but still figures in its accuracy conditions. Given this, the idea that the self is implicit in perceptual representation entails that perceptual representations lack constituent representations of self, although the self figures in their accuracy conditions.

While this view that the self is implicit correctly describes the most basic perceptual representations, I will argue that it does not hold for more complex, but still low-level, perceptual representations of space. I propose that in the course of percep-

[1] The idea that the self figures in perception implicitly echoes theories of self-consciousness that treat it as a peripheral, albeit necessary, feature of consciousness. For example, Kriegel (2004) argues that 'it is impossible to think or experience something consciously without thinking or experiencing it self-consciously, i.e. without being peripherally aware of thinking or experiencing it' (p. 200). Here, peripheral awareness of a thought or experience contrasts with focal awareness of its object. Similarly to the view that the self plays an implicit rather than explicit role in perceptual representation, Kriegel takes self-consciousness to be constituted by peripheral rather than focal awareness. Both views – of perceptual representation on the one hand and conscious experience on the other – thus involve a subjective element that operates 'in the background.'

[2] The view frequently expressed in the literature holds that the self is implicit in personal level perceptual representations. In §2, I'll argue that we can interpret this view as applying to sub-personal perceptual representations as well.

[3] My use of the term *constituent* differs from its use in Perry (1986). For Perry, constituents are the objects that propositions are about (p. 139). Constituents, in this sense, can be articulated or unarticulated depending on whether they are designated by the components of a proposition. For example, when I utter 'It is raining' in New York, rain is an articulated constituent of my utterance's content because it is designated by the component 'raining,' while New York is an unarticulated constituent of the content because it is not designated by any of the utterance's components. A *component*, as Perry uses the term, seems to be an element of a sentence. My use of the term *constituent* roughly maps onto Perry's use of the term *component* insofar as I use *constituent* to designate the elements of mental representations.

tual processing, implicit information about the self is made explicit. My discussion centers on an account of egocentric reference frames that describes their content and use in perception. Egocentric reference frames, which are standard posits in both psychological and philosophical theories of perception, are frameworks for specifying spatial relations relative to a perceiver. If, as I suggest, these reference frames carry information about the self, then whether such information is implicit or explicit in perceptual representations is a matter of whether egocentric reference frames are constituents of those representations. I argue that this, in turn, depends on an egocentric reference frame's role in perception. When an egocentric reference frame is used to specify the locations of other objects, it is not explicitly represented. But in the next stage of perceptual processing, an egocentric reference frame's role changes: it serves as the object whose own location is specified relative to another frame. In its new role, the frame is explicitly represented. In this way, implicit information about the self becomes explicit in the next stage of processing. This progression, where implicit information at one stage becomes explicit in the next, continues, forming a hierarchy of reference frames.

I proceed as follows. In §2, I argue that information about the self is implicit in perceptual representations in which an egocentric reference frame is used to specify the locations of other objects. In §3, I propose that this implicit information becomes explicitly represented as perceptual processing advances and one frame is explicitly represented at a location relative to another. My proposal rests on a view I develop called the *Nested Frames View*, which posits that whether egocentric reference frames are implicit or explicit in perceptual representation depends on their role in a hierarchy of nested frames. §4 shows how standard explanations of two perceptual phenomena – gaze shifts and vection, the illusory experience of self-motion – are committed to and thus support the *Nested Frames View*. I briefly conclude in §5.

2 Implicit Information about the Self in Perception

This section articulates the view that information about the self is implicit in perceptual representation and sketches two of its standard motivations. I then find further support for this view by analyzing egocentric reference frames and their role in perception; however, this support is limited. My analysis suggests that information about the self is implicit in only the most basic perceptual representations. An updated view appears needed to adequately describe all perceptual representations.

To begin, let the view that information about the self is implicit in perceptual representation be articulated as follows:

Implicit Self: Perceptual representations do not include constituent representations of self, although the self figures in the accuracy conditions of perceptual representations.

Suppose I visually perceive a book three feet in front of me. According to *Implicit Self*, my complex visual representation does not include a constituent representation of self. The constituent representations are of an object (the book), a distance (three feet), and a direction (in front of), but there is no constituent representation of my self.[4] Despite this, I figure in the accuracy conditions of my visual representation. My visual representation is accurate just in case a book is three feet in front of *me*. *Implicit Self* holds that this is true for all perceptual representations: they lack constituent representations of self, although the self figures in their accuracy conditions.[5]

My interest here is in how information about the self figures in perceptual representations generally, but this presents a slight disconnect with the literature, which typically focuses on personal level perceptual representations. The view more frequently expressed in the literature can be glossed as holding that personal level perceptual representations lack constituent representations of self, although the self figures in their accuracy conditions (Ismael 2012, p. 72; Musholt 2015, pp. 80–81; Recanati 2012, pp. 185–190). I believe that if you are committed to this, you are likewise committed to the view that information about the self is implicit in all perceptual representations. This is because the arguments that support the former also support the latter. These general arguments – one of which concerns simplicity and the other, immunity to error through misidentification – are discussed below. In any case, my focus going forward is on perceptual representations in general, and to the extent that *Implicit Self* correctly describes personal level perceptual representations, I take it to likewise correctly describe sub-personal level perceptual representations.

One theoretical motivation for *Implicit Self* is based on considerations of simplicity. One begins with the observation that, regardless of how the scene that one perceives changes, one's self is *always* the subject of perception. Because there is no variation in the role of the self in perception, it seems that a perceptual representation of self is not needed. Representations, it is assumed, are only needed to track things that are variable. Occam's Razor does the rest: because a representation of self is not needed, no such representation is used. A representation of self would be superfluous.

Implicit Self also allows us to explain why certain judgments of the form '*I am F*' are immune to error through misidentification (IEM). A judgment of this form is susceptible to error through misidentification if one's grounds for the judgment are such that it is possible for one to know that something is F and be mistaken that the thing that is F is oneself (Evans 1982; Shoemaker 1968). To borrow from Shoemaker (1968), the thought '*I am bleeding*' is susceptible to error through misidentification when based on visual observation (p. 556). I might know that someone is bleeding but misidentify the person bleeding as myself when it is in fact my friend. Judgments that are IEM are not susceptible to this kind of error. For example, the thought '*I have a toothache'* is IEM when based on an inner feeling of pain: I cannot know that someone has a toothache and be mistaken that the relevant person is myself (Wittgenstein 1958, pp. 66–67).

A popular method used to explain why certain thoughts are IEM is to posit that their grounds never include identifications of the form '*I* = *a*'. If the grounds of a *de se* thought lack an identification of this form, then 'it looks as though there is no point

[4] To clarify, I take it that I perceptually represent a distance of three feet, but not that I represent the book's distance from me in units. After all, it would seem arbitrary for me to specify the book's distance from me in feet rather than meters. The same considerations apply to how I perceptually represent directions. See Peacocke (1992, p. 69) for discussion of the idea that spatial perception is unit-free.

[5] The scope of *Implicit Self* is limited to cases in which the self is only the subject of perception. The thesis does not describe perceptual representations in which one is also an object of perception, such as when one looks in a mirror and sees oneself.

at which an error of *mis*identification could get introduced' (Morgan and Salje 2020, pp. 154–155). By contrast, if the grounds of a *de se* thought do depend on an identification of this form, then there is the possibility of misidentifying oneself. But how can the grounds of a *de se* thought lack an identification of the form '$I = a$'? Ismael (2012), Musholt (2015), and Recanati (2007, 2012) propose that non-conceptual representations of the form '*F*' are directly translated into conceptual representations of the form '*I am F*', bypassing the need for such an identification. Critically, this entails that non-conceptual representations, like perceptual representations, lack constituent representations of self.[6]

Having outlined these two motivations for the idea that information about the self is implicit in perceptual representation, I turn to offer a new argument in its defense. In the remainder of the section, I present an account of egocentric reference frames, describing their contents and use in perception, and argue that this account provides limited support for the claim that information about the self is implicit, rather than explicit, in perceptual representation.

To begin, let's consider the role that has been attributed to egocentric reference frames in the psychology of perception and draw out key lessons about the contents of perception. Psychologists posit egocentric reference frames to explain how we specify the locations of objects relative to ourselves (Colby and Duhamel 1996; Gallistel 1990; Groh 2014; Julesz 1971). The thought is that when, for example, I see a book in front of me or hear music behind me, my perceptual representation specifies the location of the object perceived relative to an egocentric reference frame. But discussions of these reference frames often leave open the nature of the 'ego' at the center of the theory. To address this, I'll distinguish two functions of egocentric reference frames and argue that these functions motivate a semantic account according to which these frames carry information about the self.

First, egocentric reference frames function as spatial reference frames. They specify the locations of objects using a coordinate system whose origin corresponds to a spatial point and whose axes correspond to a set of directions (Evans 1982, pp. 153–154; Peacocke 1992, ch. 3). Creatures with bodily joints often use many egocentric reference frames, each with a different origin and set of axes, to specify object locations relative to their different body parts. For example, the locations of visually perceived objects are thought to be mapped relative to the eyes in a *cyclopean reference frame*, whose origin corresponds to the perceiver's cyclopean eye (i.e., the midpoint between the eyes) and whose axes correspond to directions relative to that point (Julesz 1971; Ono et al. 2002). A *body-centered* (or trunk-centered) *reference frame*, by contrast, maps objects relative to the torso. Its 'origin is given by the property of being the center of the chest of the human body, with the three axes given by the directions back/front, left/right, and up/down with respect to that center' (Peacocke 1992, p. 62). Insofar as they are used to specify object locations, egocentric reference frames resemble allocentric reference frames. While allocentric frames are character-

[6] I take it that perceptual representations are non-conceptual (see Block 2023). Following Evans (1982), I gloss this as meaning that a perceiver might token perceptual representations without possessing the concepts necessary to specify the contents of those representations. (I leave the debate regarding the distinction between state and content non-conceptualism aside (see Bermúdez 2007; Heck 2000; Toribio 2008).)

ized in various ways – sometimes as 'centred on a point in space distinct from the one that the perceiver is occupying' (Schellenberg 2007, p. 614) and sometimes as 'non-centred structured sets of spatial relations' (Fernandez Velasco 2024, p. 164) – they too function to specify object locations.

Second, egocentric reference frames imbue spatial representations with the practical significance required to motivate action. To borrow from Burge (2011), such frames treat their origins as 'having immediate ego-relevance' (p. 294).[7] In this respect, they differ from allocentric reference frames. To see this, consider the parallel contrast between *de se* and *de dicto* thoughts. The thought *'my pants are on fire'* will prompt protective action because it is *de se*. By contrast, the *de dicto* thought *'those pants are on fire'* – even when unwittingly thought about one's own pants – will not prompt action unless connected to additional *de se* information (Kaplan 1989, p. 533; see also Perry 1979). Similarly, a perceptual representation of a baseball approaching the origin of an egocentric reference frame will prompt evasive action, while a perceptual representation of a baseball approaching the origin of an allocentric reference frame will not, unless combined with additional egocentric information. The egocentric reference frame thus seems to give information about the location of the baseball the practical significance needed to motivate action.[8]

I want to propose that these two functions motivate a view of egocentric reference frames according to which they carry two kinds of indexical information: indexical information about a location and indexical information about the self. The former fixes the origin of a spatial reference frame; the latter gives the frame practical significance by specifying whose it is. In other words, locational information is what makes something a reference frame, and information about the self is what makes a reference frame egocentric.[9] More precisely, we can think of an egocentric reference frame as a special kind of indexical that picks out a pair – a location and an individual – in the context of use. These are related by a constraint that specifies which point on the individual's body occupies the relevant location. This constraint varies across frames. For my cyclopean reference frame, for instance, the location referred to must

[7] Burge (2011) takes what he calls an 'egocentric index' to have two functions that closely parallel the two I have described. One function is 'to index an origin for a framework of representation' (p. 294). While this overlooks the role of axes, the idea is that an egocentric index functions to structure a reference frame. Its other function, Burge writes, is to 'type-individuate an aspect of psychological states that treats the origin as having immediate ego-relevance' (p. 294). Among other things, the notion of ego-relevance captures relevance to one's 'own needs, goals and perspective' (p. 294). This seems broader than the function of egocentric reference frames to help motivate action. I have chosen to focus on the latter, specifically, given its close parallel to analyses of *de se* thought.

[8] That perceptual representations must have the practical significance required to motivate action is often used as a reason to posit that they are *de se* (Alsmith 2017; Bermúdez 1998; Brewer 1992; Hurley 1998). This idea has been challenged by those who argue that perceptual representations can have merely indexical locational contents, rather than *de se* contents, and still give one reason to act (Mitchell 2021; Peacocke 2016; Schellenberg 2016).

[9] This proposal relates to the idea that phenomenal consciousness essentially involves *for-me-ness*. As Zahavi and Kriegel (2015) explain, when I am in a phenomenal state, there is not merely something it is like to be in that state; rather, there is something it is like *for me* to be in that state. In this way, all my experiences share the feature of being *for me*. The suggestion that egocentric reference frames contribute self-related information to perceptual content may help illuminate how perceptual experiences have this subjective feature.

be occupied by the midpoint between my eyes. By contrast, for my body-centered reference frame, the location referred to must be occupied by the center of my chest.

These two kinds of content are jointly constitutive of egocentric reference frames. Without locational information, an egocentric reference frame cannot serve as a spatial reference frame; without information about the self, it cannot confer the practical significance needed to motivate action. This distinguishes egocentric frames from locational and *de se* indexicals in thought and language. Unlike egocentric reference frames, these other indexicals can be used independently to express either locational or self-related information.

Crucially, because an egocentric reference frame is constituted by these two contents, they always remain packaged together. If the frame is explicitly represented, both its locational and self-related information are explicit; if it is implicit, both are implicit. This marks another difference between egocentric reference frames and locational and *de se* representations in thought and language. The thought *'I am sitting'*, for example, has an explicit *de se* component and implicit locational information. The contents of egocentric reference frames, by contrast, cannot be teased apart for separate implicit and explicit representation.

The semantic account that I have proposed shows how egocentric reference frames might supply perceptual representations with information about the self. But why should we believe that this information is implicit in perceptual representation, as *Implicit Self* holds? I contend that we have reason to think this because of the invariance in how egocentric reference frames are used in perceptual processing. Different egocentric reference frames appear to be preferentially implemented in different brain regions (see Andersen et al. 1985; Barendregt et al. 2015; Gross and Graziano 1995; O'Keefe and Dostrovsky 1971). Since perceptual inputs reach these areas in a fixed temporal order, the sequence in which inputs are used with different reference frames appears fixed. That is perceived objects are represented at locations first relative to one egocentric reference frame, and then a second, and then a third, and so forth. And this order does not seem to change from one perceptual representation to the next.

The invariant use of egocentric reference frames in perceptual processing suggests that they are best understood as encoded in the functional architecture of our perceptual systems. Functional architecture roughly amounts to the 'hardware' governing how the representations used by a particular system are processed. Pylyshyn (1984) explains that we have a choice between attributing mental functions to functional architecture or representations when developing cognitive models. He further advises that, when faced with this choice, one of our goals should be to develop models that 'fix as many properties as possible by building them into the fixed, functional architecture' (p. 106). Put differently, our methodological approach should involve attributing mental functions to functional architecture, rather than representations. Doing so constrains a model's expressive power and thereby more closely tailors the model to its explananda. If we accept Pylyshyn's advice, it seems that we should interpret egocentric reference frames as built into the fixed, functional architecture of our perceptual systems. Doing so seems to provide a 'principled rationale' for why the use of such reference frames appears invariant at certain stages of perceptual processing (p. 106). While the idea of content being architecturally encoded awaits further analysis, its application here is clear enough. Interpreting my cyclopean reference frame as architecturally encoded, for example, seems to explain why early visual representations are always used with this frame and not another: my visual system follows a functional rule regarding the use of my cyclopean reference frame in early visual processing, and as such, use of the frame at this stage of processing is invariant.

The claim that egocentric reference frames are architecturally encoded supports *Implicit Self*, at least for a given stage of perceptual processing. This is because if egocentric reference frames are architecturally encoded, then we have no reason to posit that they are also explicitly represented. Thus, their content – information about the self – must be implicit.

I have suggested that egocentric reference frames are implicit in perceptual representation. This describes them in their customary role: when used as the reference frame relative to which objects are represented. But perception also involves specifying the locations of egocentric reference frames relative to each other. In cases like these, one reference frame takes on the role of the object whose location is specified. To account for their dual role, I turn to present the *Nested Frames View* as a new account of how egocentric reference frames are used in perception. According to the view, egocentric reference frames are implicit in perceptual representation when they serve as the reference frame relative to which other objects are represented, but explicit when they play the role of an object. Granting that egocentric reference frames carry information about the self, this entails that some perceptual representations carry explicit information about the self. In §4, I then review cases that support the *Nested Frames View*, and provide evidence, contra *Implicit Self*, for the restricted use of representations of self in perception.

3 The *Nested Frames View*

In the previous section, I argued that *Implicit Self* captures an important aspect of perception. It describes perceptual representations in which the locations of objects are specified relative to an egocentric reference frame. I contended that in basic representations like these, we have reason to believe that the frame used is implicit, and consequently, that the information it carries about the self is also implicit. However, we should be cautious about accepting *Implicit Self* as a general principle. This skepticism arises because egocentric reference frames are not limited to functioning as reference frames, as they do in standard cases; they also seem to function as objects whose locations are specified. This occurs in creatures like us, whose perceptual systems use many egocentric reference frames and must track their variable spatial relations. When one reference frame is represented at a location relative to another, the former assumes the role of an object while the latter serves as a reference frame. To account for these cases, this section proposes the *Nested Frames View* as a new theory of how egocentric reference frames are used in perception. As I will discuss, the view seems to show how the information about the self carried by such frames comes to be explicitly represented in perception.

For context, empirical research on reference frames has often focused on where in the brain different frames are implemented and what cognitive processes they underwrite (Andersen and Zipser 1988; Barendregt et al. 2015; Chen et al. 2013; Colby and Duhamel 1996; Groh 2014; O'Keefe and Dostrovsky 1971). In philosophy, a central concern has been why our experiences appear unified despite the use of many frames (Alsmith 2020; Briscoe 2009; Grush 2000). My question here is different: how are reference frames represented relative to one another, and what does this reveal about how the self is represented in perception?

The *Nested Frames View* addresses this question in two parts. The first concerns the representational relations between the egocentric reference frames used by a creature in perception.[10] The view proposes that within a given perceptual system, one reference frame is nested within another, meaning that the former is represented at a location relative to the latter. This nesting relation then iterates. Imagine a creature that uses three reference frames: RF_0, RF_1, RF_2. If RF_0 is nested within RF_1 and RF_1 is nested within RF_2, this means that RF_0 is represented at a location relative to RF_1 and RF_1 is represented at a location relative to RF_2.[11]

The second component of the *Nested Frames View* describes when egocentric reference frames are implicit and explicit based on their role in the nested structure. This builds on insights from existing work suggesting that egocentric frames can shift between being implicit and explicit. Grush (2000), for instance, argues that what he calls a point of view (POV) functions as the implicit anchor of an egocentric space but becomes explicitly represented when coordinated with an allocentric frame (p. 82). While Grush focuses on egocentric-to-allocentric coordination, the *Nested Frames View* extends his idea to egocentric-to-egocentric relations. Specifically, the view holds that for any given pair of nested frames, the nested frame is explicit and the nesting frame is implicit. In other words, when one frame is represented at a location relative to another, the former is explicitly represented and the latter is architecturally encoded. This entails that when RF_0 is nested within RF_1, which is nested within RF_2, RF_0 is explicitly represented at a location relative to RF_1, which is architecturally encoded, and RF_1 is explicitly represented at a location relative to RF_2, which is architecturally encoded.

To illustrate this nested structure more concretely, let's look at an example. Imagine a creature whose visual system uses four egocentric reference frames that are arranged such that RF_0 is nested within RF_1, RF_1 is nested within RF_2, and RF_2 is nested within RF_3. I will stipulate that perceived objects are first represented at locations relative to the maximally nested reference frame, RF_0; the reason for this will become clear later. When perceived objects are explicitly represented at locations relative to RF_0, it remains architecturally encoded. But when its location is specified relative to RF_1, RF_0 is explicitly represented. This pattern, in which one reference frame is architecturally

[10] The *Nested Frames View* seems to capture the representational relations between the egocentric reference frames used in human visual perception, as will be seen in §4. I anticipate that the view also applies to human auditory perception and similar perceptual systems in non-human animals. However, how far the view extends beyond that remains an open question.

[11] Other configurations are also compatible with the view. For instance, both RF_0 and RF_1 could be nested within RF_2. This might occur in vision, if two monocular reference frames are nested within the cyclopean reference frame.

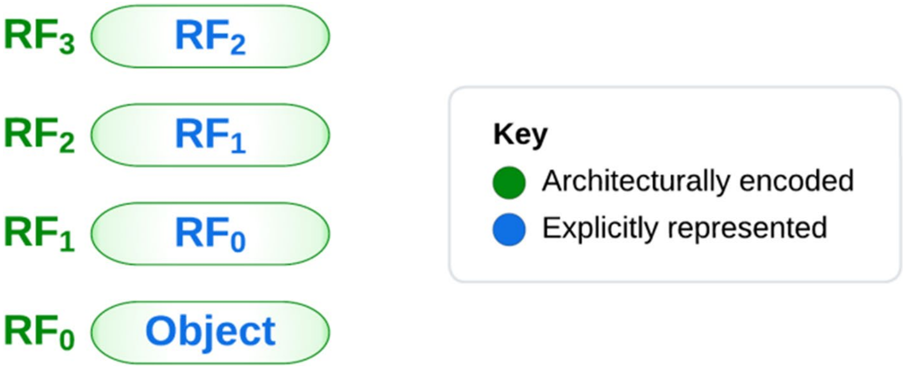

Fig. 1 Reference frames in a nested structure. Perceived objects are nested within RF_0, RF_0 is nested within RF_1, RF_1 is nested within RF_2, and RF_2 is nested within RF_3

encoded at one level of the nested structure and explicitly represented in the next, continues, as shown in Fig. 1. The pattern stops with the outermost frame, RF_3, which is not explicitly represented because it is not nested within another frame.

Now, let me outline what the nested structure used in human visual perception might look like. For simplicity, let us assume that only four reference frames are used: the cyclopean reference frame, the head-centered reference frame, the body-centered reference frame, and an allocentric reference frame.[12] We should expect that the cyclopean frame is nested within the head-centered frame, which is nested within the body-centered frame, which is nested within an allocentric frame.[13] This is because objects in the environment that are visually perceived are represented relative to the eyes before the head; they are represented relative to the head before the body; and they are represented relative to the body before the environment. In this way, maximally embedded egocentric reference frames are located at the beginning of the visual processing stream; but as processing continues, information is passed to frames that occupy less and less embedded positions. In fact, moving along the processing path critically involves contextualizing incoming information in wider and wider spatial frameworks.

I now turn to consider three lines of support for the *Nested Frames View*: empirical findings from spatial neglect, robot kinematics, and an analysis of proprioceptive information. Together, these reinforce not only the general idea of a hierarchy of egocentric reference frames in perception but also the specific structure outlined above for human visual perception.

[12] This structure leaves out egocentric reference frames, like monocular reference frames, whose role in perception is less well-established. It also does not engage with the possibility of more complex kinds of reference frames, such as hybrid reference frames (Carrozzo and Lacquaniti 1994) and idiosyncratic reference frames (Chang and Snyder 2010).

[13] The hierarchical arrangement of reference frames is an empirical matter and will vary across systems. For example, Nortmann et al. (2025) developed a recurrent neural network that converts egocentric (eye-centered) coordinates into allocentric (image-centered) coordinates while minimizing energy consumption. Because this model uses only two frames, it cannot implement a hierarchy like the one I propose for human visual perception.

First, studies of spatial neglect reveal error patterns consistent with a hierarchy of reference frames in perception. Before considering the empirical findings, it is helpful to clarify how errors would propagate within such a hierarchy. Given that information flows upward, errors in lower-level reference frames should propagate upward to higher-level frames, but not downward. For instance, using Fig. 1, an error in RF_2 should produce errors in RF_3, while leaving RF_0 and RF_1 unaffected. Applied to the hierarchy proposed for human visual perception, errors in the body-centered frame should affect allocentric representations but not head-centered or cyclopean representations.

Empirical evidence supports this prediction. Li et al. (2014) studied patients with left-sided spatial neglect using triangular targets that either contained a gap or were intact. The targets appeared at different positions relative to the patient's trunk, while the gap's location varied in allocentric space. Crucially, neglect of gaps on the left of allocentric space was worse when stimuli appeared to the left of the trunk. The authors concluded that 'not only were allocentric and egocentric biases present simultaneously, but that egocentric information can influence the severity of allocentric neglect' (p. 166).[14] Focused on disambiguating between different egocentric reference frames, Karnath et al. (1991) also studied patients with left-sided spatial neglect and found that it was specifically trunk-centered: leftward trunk movements compensated for neglect, while head movements did not. This suggested that the patients' spatial neglect was localized to their body-centered frame and did not affect their head-centered frame (see also Karnath et al. (1993)).

Taken together, these findings provide empirical grounds for a hierarchical organization of egocentric reference frames used in visual perception. The influence of body-centered errors on allocentric representations, alongside the absence of errors in the head-centered frame, suggests a structure in which the body-centered frame falls between the lower-level head-centered frame and a higher-level allocentric frame. By contrast, a non-hierarchical model appears incompatible with these results. If, for example, the body-centered reference frame were explicitly represented relative to both an allocentric frame and the head-centered frame, it would be difficult to explain why errors in the body-centered frame do not propagate to the head-centered frame.

That said, it is important to recognize that a hierarchical model does not preclude top-down influences from higher-level frames on lower-level ones. Increasing evidence shows that hippocampal information, which encodes an allocentric (or cognitive) map, can modulate early visual processing. As Fernandez Velasco (2024) emphasizes, 'The classical picture (be it in neuroscience, psychology, or philosophy) of the interaction between the systems in charge of visual processing and spatial location is bottom-up… A wealth of discoveries is now turning this classical picture on its head' (p. 162). For example, recent studies have demonstrated that V1 neural responses to landmarks are modulated by allocentric information about self-location (Saleem et al. 2018) and that some V1 neurons exhibit stimulus-predictive responses possibly 'scaffolded' by hippocampal activity (Fiser et al. 2016, p. 1664). While these

[14] A hierarchical relationship between egocentric and allocentric processing has been observed at the neural level by Zaehle et al. (2007), who suggest that 'a hierarchically organized processing system exists in which egocentric spatial coding requires only a subsystem of the processing resources of the allocentric condition' (p. 92).

findings reveal top-down effects on early visual processing, they do not constitute evidence that higher-level frames systematically alter the spatial positions of objects represented in lower-level frames, which would suggest that reference frames are mutually represented relative to each other rather than nested. Thus, the current findings appear compatible with the *Nested Frames View*.

A second source of support for a hierarchy of reference frames in perception comes from robotics (see standard robotics textbooks such as Craig (2009) and Siciliano et al. (2008)). In robot kinematics, each of a robot's links or articulated body parts is assigned its own reference frame. These frames are then arranged hierarchically, with each defined relative to a parent frame. The root of the hierarchy, known as the 'base frame,' is typically anchored to the most stable part of the system (e.g. the torso in humanoid robots). Other frames represent links beyond the base that can move given the robot's joints. Movement is modeled by specifying the spatial relations between each frame and its parent, tracing all the way back to the base frame.

Like robotic systems, biological systems might also use a hierarchy of reference frames because they too must track a hierarchy of physical relations introduced by joints. Consider the relationship between the eyes and head. Physically, the eyes can move while the head remains fixed, but the head cannot move while keeping the absolute position of the eyes fixed. In keeping with computational models in robotics, the nested frames structure I have proposed for visual perception suggests that this asymmetry is mirrored at the representational level: the cyclopean frame can be represented at different positions within the architecturally encoded head-centered frame, but not vice versa. Like the head relative to the eyes, the head-centered reference frame functions as the anchor relative to which the cyclopean reference frame is represented. Given that the asymmetrical relations between body parts continue (e.g. the head moves relative to the torso), we should expect the asymmetrical relations between reference frames to continue, thereby forming a hierarchy.

The use of hierarchical spatial structures like those found in robotics has precedent in other models of perception. Bermúdez (2017) explicitly draws on robotics in developing a model of bodily awareness. He proposes that the human body is represented 'as a hierarchy of generalized cones linked by mechanical joints' and anchored to the immovable torso (p. 136). According to his view, a sensation in the hand, for example, is located in a cylindrical coordinate system centered on the wrist that is then located relative to the torso via joint angles at the elbow and shoulder. In this way, a wrist-centered frame is nested within an elbow-centered frame, which is nested within a shoulder-centered frame, which is nested within the torso-centered frame. While Bermúdez uses this kind of hierarchical structure to model spatial representation in interoceptive perception, the *Nested Frames View* uses it to model spatial representation in exteroceptive perception.

The third and final source of support for the *Nested Frames View* that I will discuss concerns the proprioceptive information used to track spatial relations between reference frames. I'll explicate my idea by previewing an example discussed in §4. During gaze shifts, cognition must track object positions relative to the eyes and eye position relative to the head in order to represent visual objects relative to the head (Andersen et al. 1993, pp. 171–173; Briscoe 2021, p. S3926; Grush 2000, p. 68; Zipser & Andersen 1988). Mathematically, representing eye position relative to the head is equivalent to representing head position relative to the eyes. But when we consider the source of this spatial information, a clear asymmetry emerges, one closely related to the asymmetry between body parts. The information used is proprioceptive and derives from eye muscle activity during eye movements (Balslev and Miall 2008; Briscoe 2021, p. S3926; Zipser & Andersen 1988). Thus, this information is best construed as concerning the position of the eyes relative to the head rather than vice versa. This indicates a possible asymmetry between the relevant reference frames, namely, that the cyclopean reference frame is represented at a location relative to the head-centered frame rather than vice versa. More broadly, this suggests a hierarchy of reference frames in which each frame is represented relative to another, mirroring the asymmetric proprioceptive signals that track the spatial relations between body parts.

Beyond providing evidence for a hierarchy, this analysis of proprioceptive information helps highlight a key feature of the *Nested Frames View*: that some reference frames are explicitly represented, in addition to being architecturally encoded. This likely occurs because our perceptual systems must track the variable relations between different reference frames. We've seen, for example, how representing visual objects relative to the head requires tracking both their positions relative to the eyes and the eyes' position relative to the head. This suggests that the cyclopean reference frame plays two roles: it serves as a frame relative to which objects are represented and as an object represented relative to the head-centered frame. When objects are represented at locations relative to the cyclopean frame, it remains architecturally encoded. As we saw in §2, there is no need for it to be represented. Yet when the eyes shift, downstream parts of cognition must track and compensate for this shift. I propose that at one step higher in the processing stream, the cyclopean frame is now explicitly represented at a location relative to the architecturally encoded head-centered frame. Thus, the cyclopean frame is encoded twice over: first, implicitly, for representing other objects at locations relative to it; second, explicitly, for representing itself relative to another frame. This reflects a broader functional trend: an implicit egocentric reference frame functions as the frame relative to which other objects are represented, while an explicit egocentric reference frame functions as an object that is itself represented relative to another frame.

The dual use of egocentric reference frames mirrors the self's roles as subject and object of perception. To see this, notice how the functions of the cyclopean reference frame map onto the two roles of the eyes: to perceive and to be perceived. When the cyclopean reference frame is architecturally encoded and used as the frame relative to which objects are represented, it carries information about the eyes in their perceiving role. The eyes are perceiving the objects represented at locations relative to the cyclopean frame. By contrast, when the cyclopean reference frame is explicitly represented at a location relative to another frame, it carries information about the eyes as they are being perceived. In this case, the eyes are objects of proprioception, whose positions relative to the head must be tracked for downstream perceptual processing. Interestingly, even while reflecting the role of the eyes as objects of proprioception, the cyclopean reference frame serves as an input to downstream perceptual processing and so facilitates one's role as a perceiver. In this way, as a whole, the nested structure of reference frames is marshaled in support of the self's role as the subject of perception.

The nested structure of egocentric reference frames illustrates how indexical information about the self comes to be explicitly represented in perception. When an egocentric reference frame that is architecturally encoded at one level of the nested structure becomes explicitly represented in the next, the information that it carries is likewise articulated. This follows from my analysis of egocentric reference frames as structures that jointly encode indexical locational information and indexical information about the self (§2).

Another source of support for the idea that explicit representation of an egocentric reference frame entails explicit representation of information about the self comes from consideration of conscious access. The positions of our body parts relative to one another are consciously accessible (Bermúdez 1998, ch. 6; Evans 1982, ch. 7). I can, for example, feel where my eyes are located relative to my head and report on this. Crucially, this feeling is accompanied by a sense of ownership (de Vignemont 2018). When I feel the position of my eyes relative to my head, they feel like *my* eyes. That something is consciously accessible in this way is generally taken to be indicative of explicit representation. Thus, this consciously accessible sense of ownership provides evidence that when information about eye position is explicitly represented, information about the self – here, information that the eyes are *my* eyes – is also explicitly represented. More generally, this suggests that information about the self is explicitly represented when egocentric reference frames are so represented.

We can use the nested structure of reference frames proposed earlier to sketch the use of explicit information about the self in human visual perception. The cyclopean reference frame's nesting within the head-centered frame marks explicit information about one's eyes, used to specify their position relative to the head. The head-centered reference frame's nesting within the body-centered frame marks explicit information about one's head, used to specify its position relative to the body. And the body-centered reference frame's nesting within an allocentric frame marks explicit information about one's body, used to specify its position relative to the environment.

In summary, the *Nested Frames View* proposes that egocentric reference frames are hierarchically organized such that they can be both architecturally encoded and explicitly represented depending on their role. I have provided support for this hierarchical organization using studies of spatial neglect, robot kinematics, and an analysis of proprioceptive information. Crucially, the view entails that when an egocentric reference frame moves from being architecturally encoded at one level to being explicitly represented at the next, the indexical information about the self that it carries likewise becomes articulated. I now turn to argue that the *Nested Frames View* is well-suited to describe two phenomena in human perception.

4 The *Nested Frames View* in Action

In this section, I further explicate the *Nested Frames View* by showing how it applies to specific perceptual phenomena: gaze shifts and vection. Gaze shifts are movements of the eyes, and vection is the illusory experience of self-motion. I'll argue that during gaze shifts, the cyclopean reference frame is architecturally encoded when used as the frame relative to which objects are represented, but explicitly represented when used as an object whose location is specified relative to the architecturally encoded head-centered frame. Similarly, I'll argue that during vection, the body-centered reference frame is architecturally encoded when used as the frame relative to which objects are represented, but explicitly represented when used as an object whose location is specified relative to an allocentric frame. These cases thus appear to show how information about the self comes to be explicitly represented in perception.

4.1 Gaze Shifts

The first case we will look at involves the routine perceptual experience of shifting your gaze. Let us try a short exercise, shown in Fig. 2. Look straight in front of your nose, such that some object x is in the center of your visual field. Call this t_1. Holding the position of your head fixed, rotate your eyes – in other words, shift the direction of your gaze – 20° to the left. Now x is in the right periphery of your visual field, although x's position relative to your head is the same as before. Call this t_2. Notice that at both times, part of your experience is as of x's position in front of your head and you judge x to be in front of your head. These facts give us reason to believe that you represent x at a position relative to your head. But how do you do this? The visual information you receive only gives x's position relative to your eyes and the position of your eyes relative to your head changes over time. The basic answer is that the capacity to represent x relative to your head depends on tracking the position of your eyes relative to your head. But as I will argue, this suggests that egocentric reference frames are themselves explicitly represented in higher frames, as the *Nested Frames View* anticipates.

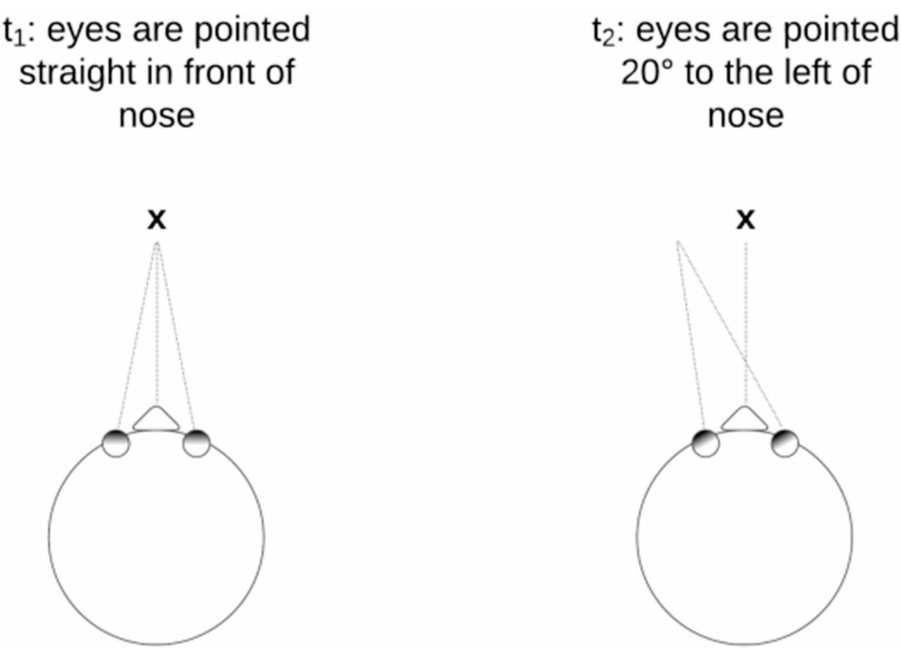

Fig. 2 Gaze shifts

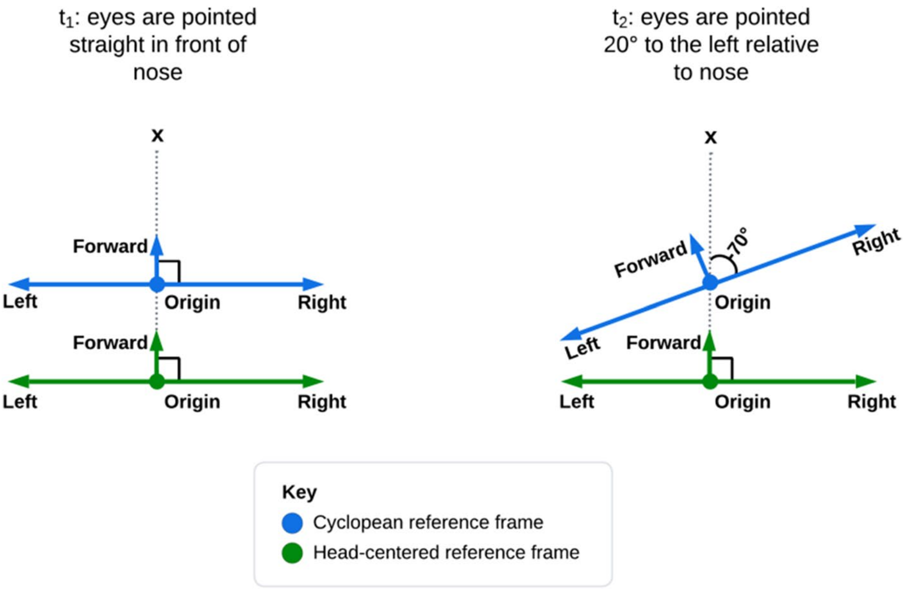

Fig. 3 A representational account of gaze shifts

During gaze shifts, the cyclopean reference frame is used to locate *x* relative to one's eyes and the head-centered reference frame is used to locate *x* relative to one's head. As shown in Fig. 3, at t_1, *x* is represented 90° from the cyclopean frame's left-right axis because *x* falls in front of one's eyes. The direction corresponding to this axis runs parallel to the direction corresponding to the left-right axis of one's head-centered frame, and so, *x* is represented 90° from the latter as well. At t_2, *x* is represented 70° from the left-right axis of one's cyclopean frame because of one's 20° gaze shift to the left. Critically, because of one's gaze shift, this axis corresponds to a direction that is 20° rotated from the direction corresponding to the left-right axis of one's head-centered frame. The 20° shift in the axis of the former frame relative to the latter offsets the 20° change in *x*'s position relative to the cyclopean frame. Thus, at t_2, *x* is still represented 90° from the left-right axis of one's head-centered frame.[15]

What this explanation of gaze shifts highlights is that *x*'s position relative to one's head-centered reference frame is a function of *x*'s position relative to one's cyclopean reference frame and the cyclopean frame's position relative to one's head-centered frame. The *Nested Frames View* accounts for this calculation by positing explicit representations of *x* and the cyclopean reference frame, both of

[15] This analysis need not be taken to imply support for what Briscoe (2021) calls the decomposition thesis, according to which 'perceptual experiences resolve without remainder into their different modality-specific components' (p. S3913). In the context of gaze shifts, this thesis would entail that awareness of an object's location relative to the head decomposes into awareness of its location relative to the eyes and awareness of the eyes' position relative to the head. Against this view, Briscoe argues that the object's location relative to the head constitutes a '*strongly novel* type of perceptible feature' that cannot be reduced to these components alone (p. S3929). While my analysis focuses on only these two components, I remain neutral on whether there are additional representations of the kind Briscoe describes.

which play the role of the object whose location is specified. Specifically, the idea is that the visual system computes *x*'s position in one's head-centered reference frame by explicitly representing *x* relative to the architecturally encoded cyclopean frame and explicitly representing the cyclopean frame relative to the architecturally encoded head-centered frame. As can be seen, this explanation relies on the idea that one's cyclopean reference frame is both architecturally encoded and explicitly represented at different stages of processing. It is architecturally encoded when perceived objects are represented at positions relative to it, and it is explicitly represented when its position is specified relative to the head-centered frame. Insofar as the cyclopean reference frame is explicitly represented relative to the head-centered frame, the former is nested within the latter.

The foregoing account posits that the visual system explicitly represents the cyclopean reference frame relative to the head-centered reference frame. The deflationary alternative is that the function used to compute the position of object *x* relative to the head-centered frame is encoded in the visual system's functional architecture. In other words, the system might follow a rule, such as: *When the eyes rotate x*$^{\circ}$ *counter-clockwise*, *the location of object o relative to the head-centered reference frame* = *the location of o in the cyclopean reference frame* + *x*$^{\circ}$. If this kind of rule is architecturally encoded, then no representation of the cyclopean reference frame is necessary. How objects are represented relative to the head-centered frame is guaranteed by how the visual system functions.

I turn to discuss two studies of visual perception, which considered together, support positing an explicit representation of the cyclopean reference frame during gaze shifts. As will be seen, the studies are framed to concern a 'representation of eye position,' but this is plausibly interpreted as a representation of the cyclopean reference frame at a location relative to the head-centered reference frame. After all, what is relevant to the studies is information about the position of one's eyes relative to one's head, and this is just the information carried by the latter representation. If this is right, then these studies give us reason to think that egocentric reference frames are sometimes explicitly represented in perception.

Wang et al. (2007) recorded neurons in area 3a of the primary somatosensory cortex of two rhesus monkeys and identified a total of 88 neurons that appeared sensitive to eye position (pp. 640–641). Given their location in the somatosensory cortex, a brain region involved in sensory processing, the researchers hypothesized – and later confirmed – that these neurons were involved in the sensory processes that inform perceptual representation. For each neuron, the strength of its neural signal correlated with a particular position of the eyes relative to the head. For example, one neuron exhibited baseline activity for all positions below the head's horizontal axis, increasing activity for positions above it, and maximum activity at 15° from its vertical axis and 0° from its horizontal axis. This neuron thus seemed tuned to an eye position of 15° from the head's vertical axis and 0° from its horizontal axis (p. 641). 70% of the neurons also showed a phasic response: when the monkey made a saccade to the neuron's preferred eye position, the signal would initially spike, before exhibiting a constant firing rate for as long as the monkey held a fixed gaze (p. 642). The correlation between neural responses and eye positions led the researchers to conclude that there is a 'representation' of eye position in the somatosensory cortex of the monkey.[16]

Wang et al. (2007) then attempted to confirm whether this eye position representation originated from proprioceptive signals from muscles around the eyes. The other possibility was that it came from efference copy – a copy of the motor commands used to direct movements of the eyes. The location of the neurons within 'a region of somatosensory cortex dedicated to muscle proprioception, [suggested] a proprioceptive origin', but to confirm this hypothesis, Wang et al. used a retrobulbar block to temporarily anesthetize and paralyze one eye (p. 642). The other eye was able to move normally. The researchers continued recording the neurons correlated with eye position and found that they stopped firing for the duration of the anesthesia and paralysis of the one eye. A signal from efference copy would be expected to continue despite the block, given that one eye continued to move freely. Thus, Wang et al. concluded that the eye position signal had proprioceptive origins, and that this signal was cut off when the eye was anesthetized.

While providing evidence of a proprioceptive representation of eye position, Wang et al. (2007)'s study does not specifically address whether this representation is used during gaze shifts. What we need is evidence that this representation is used to specify the positions of objects relative to the head. Balslev and Miall (2008)'s results indicate just this.

In Balslev and Miall (2008)'s study, human subjects were tasked with performing a 'straight-ahead' test. Participants saw an LED appear in an otherwise dark setting and instructed the experimenter to move the LED either to the left or right until it appeared straight ahead of their nose (p. 8969). In the control condition, subjects performed this task both before and immediately following fifteen minutes of low-frequency repetitive transcranial magnetic stimulation (rTMS) over their left motor cortex. (Low-frequency rTMS is used to temporarily 'turn-off' selected brain areas.) The test condition differed only in that rTMS was applied over the primary somatosensory cortex. While performance was highly accurate in the control condition, performance was inaccurate in the test condition: rTMS over the primary somatosensory cortex affected subjects' visual perceptions, shifting the visual scene to the left of their nose (p. 8970).[17] Balslev & Miall wrote, 'we interpret this shift as an error in perceived eye position' (p. 8970). Their idea was that the perceptual error in locating objects relative to the nose (i.e., head) was best understood as derived from a proprioceptive error in locating the eyes relative to the head. The error was thought to be proprioceptive given the role of the somatosensory cortex – the region affected by rTMS – in sensory processing.

One shortcoming of Balslev and Miall (2008)'s study is that rTMS occurs over a relatively large area, and thus, it could be that subjects' altered visual perceptions resulted from more generalized changes to connections both within and without the

[16] In other studies, eye position seems to modulate neural receptive fields, rather than being represented on its own (Andersen et al. 1985; Bremmer et al. 1998; Zipser & Andersen 1988).

[17] Using two different non-visual tests, the experimenters ruled out the possibility that the shift in subjects' perception of the visual scene was caused by a shift in their perception of their body mid-line (Balslev and Miall 2008, p. 8970).

somatosensory cortex. However, as Balslev & Miall note, their results match those of Wang et al. (2007). When considering these two studies together, the following picture emerges. The firing patterns of specific neurons in the somatosensory cortex correlate with positions of the eyes relative to the head. And, disruptions to these neurons correlate with inaccurate perceptual representations of object positions relative to the head. These correlations lend strong support to the idea that there is a representation of eye position which is used to locate objects relative to the head.[18] If this is correct, it seems reasonable to conclude that this representation is used during gaze shifts. What's more, if, as I have proposed, we can re-describe this representation of eye position as a representation of one's cyclopean frame relative to one's head-centered frame, then we can conclude that gaze shifts involve the use of an explicitly represented egocentric reference frame.

To recap, we aimed to explain how one is able to represent objects at positions relative to the head during gaze shifts. In line with the *Nested Frames View*, I proposed that representing an object at a position relative to the head-centered reference frame required explicitly representing the object at a position relative to the architecturally encoded cyclopean reference frame and explicitly representing the cyclopean frame at a position relative to the architecturally encoded head-centered frame. My proposal thus relied on the idea that one's cyclopean frame is nested within one's head-centered frame. I found support for this idea by looking at two empirical studies of visual perception. If, as I have argued, gaze shifts do in fact involve the nesting of one's cyclopean frame within one's head-centered frame, then they constitute a case in which, contra *Implicit Self*, information about the self is explicitly represented in perception. We have reason to believe that the cyclopean frame carries information about the self insofar as it imbues perceptual representations with the practical significance needed to motivate action (§2). When the frame is made explicit, so too is the information that it carries.

4.2 Vection

The second case that I will analyze using the *Nested Frames View* is vection, an illusion in which one inaccurately experiences oneself as moving when one is stationary (see Palmisano et al. 2015). The paradigmatic experiment for inducing vection, depicted in Fig. 4, runs as follows (see Mach 1875/2001). A subject is seated inside a drum painted with black and white vertobject at a position relative toical stripes. At t_0, both the subject and the drum are stationary and the subject experiences both as stationary. At t_1, the drum starts to rotate counterclockwise along the earth's vertical axis. At this time, the subject's experience is veridical: she experiences the drum as spinning and herself as stationary. The subject's experience then changes, typically between 2 and 20 seconds after the drum begins to spin (Riecke and Schulte-Pelkum 2013, p. 30). Although the drum is still spinning and the subject is still stationary, the subject inaccurately experiences herself to be moving clockwise relative to a station-

[18] The view that certain correlations between world states and internal states are sufficient to postulate mental representations is articulated and defended in Gallistel (1998) and Shea (2018, chap. 4).

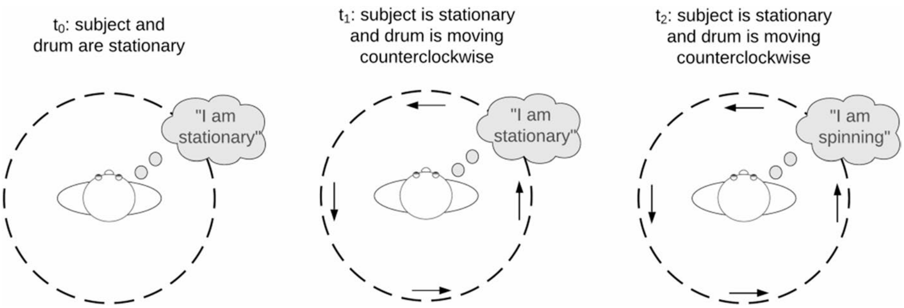

Fig. 4 Vection

ary drum. Call this t_2. I will argue that explicit representation of an egocentric reference frame offers the best representational explanation of the shift in the subject's experience between t_1 and t_2.

Vection is often understood as involving a change from one *rest frame* – the reference frame one takes to be stationary – to another (Prothero and Parker 2003; Riecke 2011; Seno et al. 2009). The idea is that one's rest frame at t_2 differs from one's rest frame at t_0 and t_1. At t_0 and t_1, when the subject takes herself to be stationary, she uses one of her egocentric reference frames as her rest frame. I will assume that this is her body-centered reference frame because the subject takes her whole body to be stationary. As shown in Fig. 5, at t_0, the drum is represented as stationary in the body-centered frame and at t_1, the drum is represented as spinning counterclockwise in this frame.

Because entire visual scenes do not normally move, at t_2, the subject's incoming visual information (of the drum moving) is interpreted as a signal that the subject is moving relative to a fixed visual scene. The visual scene – i.e. the drum – is taken as stationary, which is to say that a drum-centered reference frame is used as the subject's rest frame. Spatial relations which were previously represented relative to the subject now are represented relative to the drum. Specifically, the body-centered reference frame, which at t_0 and t_1 was used to map spatial relations relative to the center of the subject's body, is represented as moving clockwise relative to the drum-centered frame, as shown in Fig. 5.[19]

This representational account of vection comports with the *Nested Frames View*. The view suggests that at t_0 and t_1, one explicitly represents the drum relative to an architecturally encoded body-centered reference frame, while at t_2, this frame is explicitly represented relative to a drum-centered reference frame. The underlying reasons for the body-centered frame's being architecturally encoded at t_0 and t_1 but explicitly represented at t_2 concern variability and motion. At t_0 and t_1, one's body is perceived to be at rest and so, the position of one's body-centered frame is invariant. This invariance obviates any need for the frame to be explicitly represented; thus, we can posit that it is merely architecturally encoded. However, at t_2, we have

[19] Schwenkler (2014) offers a compelling *reductio* argument for the related conclusion that one's visual experience during vection involves self-locating contents. See Mitchell (2021) for a reply.

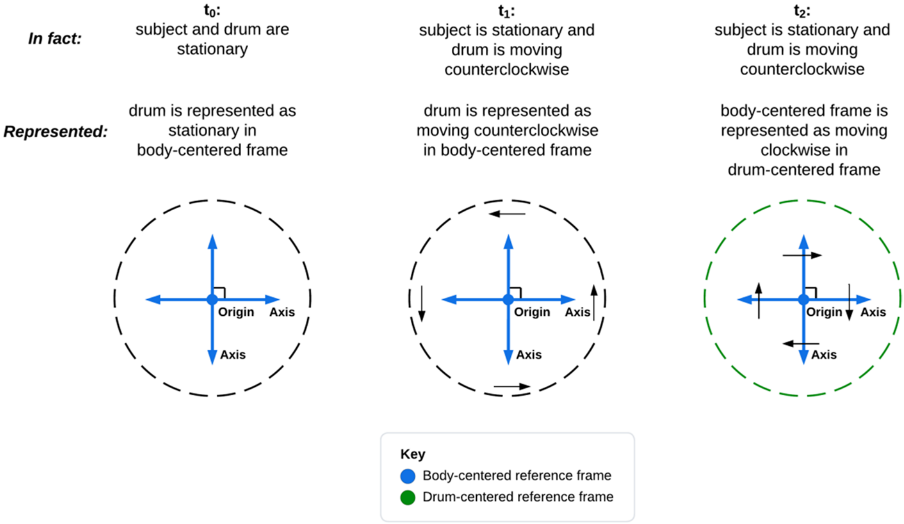

Fig. 5 A representational account of vection

reason to think that the body-centered frame is explicitly represented relative to the drum-centered frame. This is because one's body is perceived to be moving. One way to encode the body as moving is to explicitly represent it at different positions over time. Plausibly, this is achieved by explicitly representing the body-centered reference frame at different positions relative to the drum-centered reference frame.

We may reject the idea that the position of the body relative to the drum is merely architecturally encoded at t_2. This is because the spatial relations between oneself and the drum are highly variable; one can easily change both one's location and orientation relative to one's environment. Encoding such variable information in one's functional architecture would be computationally costly. For comparison, it would be far more costly than architecturally encoding the position of one's eyes relative to one's head – a deflationary proposal considered in our discussion of gaze shifts. After all, the movements of one's eyes relative to one's head are relatively constrained, while the movements of one's body in the environment are not. Given this high computational cost, we should doubt that information about one's position relative to the drum is architecturally encoded.

To sum up, I have argued that we have reason to think that during the illusory experience of self-motion, one explicitly represents one's body-centered reference frame as spinning relative to a drum-centered reference frame. Explicit representation of the former frame seems required given the body's highly variable positions relative to the drum. While the body-centered reference frame appears architecturally encoded at earlier stages of processing, I conclude that it is eventually represented explicitly. Thus, it constitutes a further example of an explicitly represented egocentric reference frame and so, a further example in which information about the self is explicitly represented in perception.

5 Conclusion

This paper has offered a new account of how the self is represented in perception. I have proposed that egocentric reference frames carry information about the self and form a hierarchy. The hierarchy reveals how implicit information about the self comes to be articulated, as each frame that is implicit at one stage of perceptual processing is explicitly represented in the next. I have marshaled empirical and theoretical considerations in support of this view, and demonstrated how standard explanations of perceptual phenomena appear committed to it. While more research is required to understand how this hierarchical organization might extend beyond perception, the tantalizing possibility is that the foundations of self-representation emerge from the ever-widening structures of egocentric reference frames used in perception.

Acknowledgments I am indebted to Gabriel Greenberg for his comments on many drafts. I would also like to thank Josh Armstrong, Katrina Elliott, Pamela Hieronymi, Jenann Ismael, Torsten Odland, Sam Pensler, Jiarui Qu, and Paul Talma for their insightful feedback and conversations, as well as audiences at the APA Central (February 2024), the International Society for the Philosophy of the Sciences of the Mind (November 2023), the Salzburg Conference for Young Analytic Philosophy (September 2023), the European Society for Philosophy and Psychology (August 2023), and the Society for Philosophy and Psychology (June 2023). Special thanks are due to my anonymous reviewers for their incisive and constructive comments.

Author Contributions Not applicable.

Funding No funding was received to assist with the preparation of this manuscript.

Data Availability Not applicable.

Materials Availability Not applicable.

Code Availability Not applicable.

Declarations

Ethics Approval Not applicable.

Informed Consent Not applicable.

Competing Interests The author has no competing interests to declare that are relevant to the content of this article.

References

Alsmith, A. 2017. Perspectival structure and agentive self-location. In F. de Vignemont, & A. J. T. Alsmith (Eds.) *The subject's matter: Self-consciousness and the body*, (pp. 263–288). MIT Press.

Alsmith, A. J. 2020. The structure of egocentric space. In F. de Vignemont, A. Serino, H. Y. Wong, & A. Farnè (Eds.) *The world at our fingertips: A multidisciplinary exploration of peripersonal space*. Oxford University Press.

Andersen, R. A., and D. Zipser. 1988. The role of the posterior parietal cortex in coordinate transformations for visual–motor integration. *Canadian Journal of Physiology and Pharmacology* 66(4):488–501.

Andersen, R.A., G.K. Essick, and R.M. Siegel. 1985. Encoding of spatial location by posterior parietal neurons. *Science* 230 (4724):456–458.

Andersen, R. A., L. H. Snyder, C.-S. Li, and B. Stricanne. 1993. Coordinate transformations in the representation of spatial information. *Current Opinion in Neurobiology* 3(2):171–176.

Balslev, D., and R. C. Miall. 2008. Eye position representation in human anterior parietal cortex. *Journal of Neuroscience* 28(36):8968–8972.

Barendregt, M., B. M. Harvey, B. Rokers, and S. O. Dumoulin. 2015. Transformation from a retinal to a cyclopean representation in human visual cortex. *Current Biology* 25(15):1982–1987.

Bermúdez, J. L. 1998. *The paradox of self-consciousness*. MIT Press.

Bermúdez, J. L. 2007. What is at stake in the debate on nonconceptual content? *Philosophical Perspectives* 21:55–72.

Bermúdez, J. L. 2017. Ownership and the space of the body. In F. de Vignemont, & A. J. T. Alsmith (Eds.) *The subject's matter: Self-consciousness and the body*, (pp. 117–144). MIT Press.

Block, N. 2023. *The border between seeing and thinking*. Oxford University Press.

Bremmer, F., A. Poutget, and K.-P. Hoffmann. 1998. Eye position encoding in the macaque posterior parietal cortex. *European Journal of Neuroscience* 10(1):153–160.

Brewer, B. 1992. Self-location and agency. *Mind* 101(401):17–34.

Briscoe, R. 2009. Egocentric spatial representation in action and perception. *Philosophy and Phenomeno logical Research* 79 (2):423–460.

Briscoe, R. E. 2021. Bodily awareness and novel multisensory features. *Synthese* 198:3913–3941.

Burge, T. 2011. Self and self-understanding. *The Journal of Philosophy* 108(6/7):287–383.

Campbell, J. 1994. *Past, space, and self*. MIT Press.

Carrozzo, M., and F. Lacquaniti. 1994. A hybrid frame of reference for visuo-manual coordination. *Neuroreport*, 5(4):453–456.

Chang, S.W., and L.H. Snyder. 2010. Idiosyncratic and systematic aspects of spatial representations in the macaque parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America* 107 (17):7951–7956.

Chen, X., G.C. DeAngelis, and D.E. Angelaki. 2013. Diverse spatial reference frames of vestibular signals in parietal cortex. *Neuron* 80 (5):1310–1321.

Colby, C. L., and J.-R. Duhamel. 1996. Spatial representations for action in parietal cortex. *Cognitive Brain Research* 5(1–2):105–115.

Craig, J. J. 2009. *Introduction to robotics: Mechanics and control, Third Edition*. Pearson Education.

Evans, G. 1982. *The varieties of reference*. (J. McDowell, Ed.). Oxford University Press.

Fernandez Velasco, P. 2024. Self-location in perceptual experience: A top-down account. *Mind & Language*, 40(2):157–173.

Fiser, A., D. Mahringer, H.K. Oyibo, A.V. Petersen, M. Leinweber, and G.B. Keller. 2016. Experience-dependent spatial expectations in mouse visual cortex. *Nature Neuroscience* 19 (12):1658–1664.

Gallistel, C. R. 1990. *The organization of learning*. MIT Press.

Gallistel, C. R. 1998. Insect navigation: Brains as symbol-processing organs. *An Invitation To Cognitive Science* 4:1–52.

Groh, J. M. 2014. *Making space: How the brain knows where things are*. Harvard University Press.

Gross, C. G., and M.S. Graziano. 1995. Multiple representations of space in the brain. *The Neuroscientist*, 1(1):43–50.

Grush, R. 2000. Self, world and space: The meaning and mechanisms of ego- and allocentric spatial representation. *Brain and Mind* 1:59–92.

Heck, R. K. 2000. Nonconceptual content and the space of reasons. *The Philosophical Review* 109(4):483–523. (originally published under the name "Richard G. Heck, Jr").

Hurley, S. L. 1998. *Consciousness in action*. Harvard University Press.

Ismael, J. 2012. Immunity to error as an artefact of transition between representational media. In S. Prosser, & F. Recanati (Eds.) *Immunity to error through misidentification: New essays*, (pp. 62–80). Cambridge University Press.

Julesz, B. 1971. *Foundations of cyclopean perception*. University of Chicago Press.

Kaplan, D. 1989. Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.) *Themes from Kaplan*, (pp. 481–563). Oxford University Press.

Karnath, H., P. Schenkel, and B. Fischer. 1991. Trunk orientation as the determining factor of the 'contralateral' deficit in the neglect syndrome and as the physical anchor of the internal representation of body orientation in space. *Brain* 114 (4):1997–2014.

Karnath, H., K. Christ, and W. Hartje. 1993. Decrease of contralateral neglect by neck muscle vibration and spatial orientation of trunk midline. *Brain* 116 (2):383–396.

Kriegel, U. 2004. Consciousness and self-consciousness. The Monist, 87(2):182–205.

Lewis, D. 1979. Attitudes *de dicto* and *de se*. *The Philosophical Review*, 88(4):513–543.

Li, D., H.-O. Karnath, and C. Rorden. 2014. Egocentric representations of space co-exist with allocentric representations: Evidence from spatial neglect. *Cortex* 58:161–169.

Mach, E. 2001. *Fundamentals of the Theory of Movement Perception*. (H. Scherberger, L.R. Young, & V. Henn, Trans.). Kluwer Academic/Plenum. (Original work published in 1875).

Mitchell, J. 2021. Self-locating content in visual experience and the 'here-replacement' account. *The Journal of Philosophy* 118(4):188–213.

Morgan, D., and L. Salje. 2020. First-person thought. *Analysis* 80(1):148–163.

Musholt, K. 2015. *Thinking about oneself: From nonconceptual content to the concept of a self*. MIT Press.

Nortmann, T., P. Sulewski, and T. C. Kietzmann. 2025. Predictive remapping and allocentric coding as con sequences of energy efficiency in recurrent neural network models of active vision. *bioRxiv*, (pp. 2025–05).

O'Keefe, J., and J. Dostrovsky. 1971. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain Research* 34(1):171–175.

Ono, H., A. P. Mapp, and I. P. Howard. 2002. The cyclopean eye in vision: The new and old data continue to hit you right between the eyes. *Vision Research* 42(10):1307–1324.

Palmisano, S., R.S. Allison, M.M. Schira, and R.J. Barry. 2015. Future challenges for vection research: Definitions, functional significance, measures, and neural bases. *Frontiers in Psychology* 6: 193.

Peacocke, C. 1992. *A study of concepts*. MIT Press.

Peacocke, C. 2016. The nature and role of first and second person content. *Analysis*, 76(3):345–354.

Perry, J. 1979. The problem of the essential indexical. *Noûs*, 13(1):3–21.

Perry, J. 1986. Thought without representation. *Proceedings of the Aristotelian Society, Supplementary volumes*, 60:137–151.

Prothero, J. D., and D. E. Parker. 2003. A unified approach to presence and motion sickness. In L. J. Hettinger, & M. W. Haas (Eds.) *Virtual and adaptive environments: Applications, implications, and human performance issues*, (pp. 47–66). CRC Press.

Pylyshyn, Z. W. 1984. *Computation and cognition: Toward a foundation for cognitive science*. MIT Press.

Recanati, F. 2007. *Perspectival thought: A plea for (moderate) relativism*. Oxford University Press.

Recanati, F. 2012. Immunity to error through misidentification: What it is and where it comes from. In S. Prosser, & F. Recanati (Eds.) *Immunity to error through misidentification: New essays*, (pp. 180–201). Cambridge University Press.

Riecke, B. E. 2011. Compelling self-motion through virtual environments without actual self-motion: Using self-motion illusions (vection) to improve user experience in VR. *Virtual Reality* 8(1):149–178.

Riecke, B. E., and J. Schulte-Pelkum. 2013. Perceptual and cognitive factors for self-motion simulation in virtual environments: How can self-motion illusions ("vection") be utilized? In F. Steinicke, Y. Visell, J. Campos, & A. Lécuyer (Eds.) *Human Walking in Virtual Environments*, (pp. 27–54). Springer.

Saleem, A.B., E.M. Diamanti, J. Fournier, K.D. Harris, and M. Carandini. 2018. Coherent encoding of subjective spatial position in visual cortex and hippocampus. *Nature* 562 (7725):124–127.

Schellenberg, S. 2007. Action and self-location in perception. *Mind* 116(463):603–632.

Schellenberg, S. 2016. *De se* content and *de hinc* content. *Analysis* 76(3):334–345.

Schwenkler, J. 2014. Vision, self-location, and the phenomenology of the 'point of view'. *Noûs* 48(1):137–155.

Seno, T., H. Ito, and S. Sunaga. 2009. The object and background hypothesis for vection. *Vision Research* 49 (24):2973–2982.

Shea, N. 2018. *Representation in cognitive science*. Oxford University Press.
Shoemaker, S. S. 1968. Self-reference and self-awareness. *The Journal of Philosophy* 65(19):555–567.
Siciliano, B., O. Khatib, and T. Kröger. 2008. *Springer handbook of robotics*. vol. 200 Springer.
Toribio, J. 2008. State versus content: The unfair trial of perceptual nonconceptualism. *Erkenntnis* 69:351–361.
de Vignemont, F. 2018. *Mind the body: An exploration of bodily self-awareness*. Oxford University Press.
Wang, X., M. Zhang, I. S. Cohen, and M. E. Goldberg. 2007. The proprioceptive representation of eye position in monkey primary somatosensory cortex. *Nature Neuroscience* 10(5):640–646.
Wittgenstein, L. 1958. *The blue and brown books*. Oxford: Blackwell.
Zaehle, T., K. Jordan, T. Wüstenberg, J. Baudewig, P. Dechent, and F.W. Mast. 2007. The neural basis of the egocentric and allocentric spatial frame of reference. *Brain Research* 1137:92–103.
Zahavi, D., and U. Kriegel. 2015. For-me-ness: What it is and what it is not. In *Philosophy of mind and phenomenology*, (pp. 36–54). Routledge.
Zipser, D., and R. A. Andersen. 1988. A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature* 331(6158):679–684.