The Hierarchy of Selves in Perception

Catherine Hochman

Abstract

The self's unique role as the subject of perception suggests that the self may have special status in perceptual representation. In accordance with this intuition, a popular philosophical view of perceptual representation holds that while the objects of perception are represented, the self is not. I argue against this view. Taking into account empirical evidence as well as philosophical considerations, I demonstrate that the self is represented in perception. In fact, many representations of self are used in perception. I propose that one's egocentric reference frames exhibit a nested structure which accounts for when and how these perceptual representations of self are used. After demonstrating how this nested structure tracks increasing degrees of agency and connects to *de se* thought, I conclude that the representations of self used in perception form a hierarchy.

1. The *Implicitness* thesis

Philosophical inquiry into questions about self-knowledge, self-consciousness, and agency seems to require an investigation of our mental states, specifically, those in which we represent ourselves. 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 a grasp of *de se* thought may be insufficient to answer our self-related questions (Bermúdez, 1998; Hurley, 1998; Peacocke, 2014). We therefore must look beyond the domain of thought and explore whether the self is represented in other kinds of mental states.

Perception provides fertile ground to examine the representation of self. 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 popular 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 usually am not among the objects I perceive, I do not perceptually represent myself. Nonetheless, what I perceive depends on where I am and so, I seem to play some implicit role in my perceptual representations. In this paper, I argue against this view that the self is implicit in perception and propose a new view – the *Nested Frames View* – as an alternative. The *Nested Frames View* holds that there are representations of self in perception and describes these representations as forming a hierarchy.

For the rest of this section, I further explicate and motivate the view that the self is implicit in perceptual representation. I also find support for this view by showing how it fits with an empirically-grounded theory of perception from perceptual psychology. In §2, I reveal that the view that the self is implicit admits of counterexamples. Specifically, I argue that in gaze shifts and vection, one's perceptual experience is best explained by positing the use of a representation of self, which constitutes explicit information about the self. In light of these counterexamples, in §3, I offer the *Nested Frames View* as a new account of when and how the self is represented in perception. The *Nested Frames View* holds that information about the self flips between being implicit and explicit. In §4, I connect the *Nested Frames View* to agency and *de se* thought to argue that the representations of self used in perception form a hierarchy. I briefly conclude in §5.

Before turning to expound the idea that the self is implicit in perception, I want to highlight two assumptions that I make about the mental representations used in perception. First, 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 possession of the concepts necessary to specify the contents of those representations.[1] This assumption is important because it widens the scope of our inquiry to include creatures usually taken to lack possession of concepts, such as infants and non-human animals.

Second, I assume that perceptual representations are *structured*: complex perceptual representations are composed of constituent perceptual representations that are bound together in a certain way. This idea can be used to elucidate the contrast between explicit and implicit invoked in the claim that the self is implicit in perception. I take information to be explicit if it is a constituent of a complex representation. For example, my complex visual representation of a red apple includes a constituent representation of red, and so, information about the color red is explicit in my visual representation. By contrast, information is implicit if it is not a constituent of a complex representation and yet it figures in the representation's accuracy conditions. To use an example from thought, when I have the thought that it is raining, my thought does not include a constituent representation of my location, and yet my thought is true just in case it is raining at my location. Thus, information about my location is implicit in my thought.

Having clarified the distinction between implicit and explicit, we can articulate the view that the self is implicit in perception as follows:

> ***Implicitness***: No perceptual representations include constituent representations of self, although the self figures in the accuracy conditions of perceptual representations.[2]

Imagine that I visually perceive an object at a location. According to *Implicitness*, my complex visual representation does not include a constituent representation of self. The constituent representations are of an object, a distance, and a direction, but there is no constituent representation of myself. Despite this, my visual representation is accurate just in case the object represented is at the distance and direction represented *relative to me*. In this way, I figure in the accuracy conditions of my visual representation. The view that the

---

[1] I leave the debate regarding the distinction between state and content non-conceptualism aside (see Bermúdez, 2007; Heck, 2000; Toribio, 2008).

[2] This view is generally formulated as concerning the *representational contents* of perceptual *experience*, but I have reformulated it to concern *all* perceptual *representations*. I focus on all perceptual representations because I take the issue of whether something is represented to come prior to the issue of whether it is represented in experience. The move from representational contents to representations should be unproblematic. This is because I understand a complex representation to have implicit representational content x if none of its constituent representations have content x.

self is implicit holds that this is true for all perceptual representations: they lack constituent representations of self, although the self figures in their accuracy conditions.[3]

A strong theoretical motivation for *Implicitness* is based on considerations of simplicity (Ismael, 2012; Musholt, 2015; Recanati, 2007, 2012; see also Perry, 1986). 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 which 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.

Implicitness 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 (see Evans, 1982; Shoemaker, 1968). To borrow from Shoemaker (1968), the thought '*I am bleeding*' is susceptible to error through misidentification when based on a 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 could not know that someone has a toothache and be mistaken that the relevant person is myself (see 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 at which an error of *mis*identification could get introduced' (Morgan & Salje, 2020, pp. 154-155). 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 lack constituent representations of self.

Given these two arguments, the view that the self is implicit in perceptual representation is well-motivated, but one may wonder how the details work out. The most pressing issue is how it is that the self figures in the accuracy conditions of perceptual representations that lack constituent representations of self. Proponents of *Implicitness* answer by invoking functional architecture, which roughly amounts to the 'hardware' governing how the representations used by a particular system are processed. Their idea is that features of the functional architecture of one's perceptual systems guarantee that the self figures in the accuracy conditions of complex representations used and processed by those systems (Musholt, 2015, p. 80; Recanati, 2007, p. 167).

While proponents of *Implicitness* do not elaborate further, we can better understand how functional architecture secures an implicit role for the self in perception by looking to perceptual psychology. A view within perceptual psychology holds that our perceptual systems function such that at certain stages of processing, the locations of objects in the environment, which are the contents of a perceptual representation, are specified

---

[3] The scope of *Implicitness* 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.

on an egocentric reference frame (Colby & Duhamel, 1996; Colby & Goldberg, 1999; Fogassi et al., 1992; Graziano et al., 1994; Gross & Graziano, 1995). An *egocentric reference frame* is a framework for specifying spatial locations by means of a coordinate system, where the origin corresponds to a point on the perceiver's body and the axes correspond to directions relative to this point (Evans, 1982, pp. 153-154; Peacocke, 1992, ch. 3). Because the use of such frames seems guaranteed, we can describe the functional architecture of our perceptual systems as encoding rules regarding the use of egocentric reference frames at different stages of perceptual processing.

When their use is architecturally encoded, egocentric reference frames are implicit in perceptual representations. In other words, they are not constituents of perceptual representations, although they figure in the accuracy conditions of perceptual representations. To see how, let us look at the *cyclopean reference frame*, which is used to specify locations relative to one's eyes (Julesz, 1971; Ono & Barbeito, 1982). The frame's origin corresponds to one's cyclopean eye (i.e. the midpoint between the eyes) and its axes correspond to directions extending from one's eyes. Again, imagine that I visually perceive an object in the environment. It is guaranteed by the functional architecture of my visual system that the object's location is specified in my cyclopean reference frame early in visual processing. In other words, in virtue of a functional rule, my complex visual representation specifies the location of the object relative to my cyclopean reference frame. My representation is accurate just in case the object represented is at the distance and direction represented relative to my cyclopean eye and directions extending from my eyes. In this way, my cyclopean reference frame appears in the accuracy conditions of my representation.

At the same time, my cyclopean reference frame is not a constituent of my visual representation. The argument for this mirrors the simplicity-based argument for *Implicitness*. Because use of the frame is guaranteed, it does not need to be represented. Given that a representation of it is not needed, considerations of simplicity suggest that such a representation is not used. This is true even though the frame's axes correspond to different sets of directions depending on the direction of my eyes. An analogy may be useful. Consider a photographer who takes pictures from different locations. We might think of the locations from which photos are taken as implicit in the photos themselves, despite that the camera's location when used to take photos varies. For our purposes: photos are to locations of the camera as early visual representations are to directions of the eyes. Despite constantly changing, the directions of the eyes remain implicit in such representations. This is just to say that the axes of the cyclopean reference frame, which correspond to directions of the eyes, remain implicit.

It is in virtue of the fact that egocentric reference frames are implicit in perceptual representations that the self is as well.[4] This is because the architecturally encoded use of such frames guarantees that the self figures in the accuracy conditions of perceptual representations. As we saw, when an early visual representation specifies an object's location in my cyclopean reference frame, the representation is accurate just in case the object represented is at the distance and direction represented relative to that which corresponds to my cyclopean reference frame. Critically, my self is part of that which my cyclopean reference frame corresponds to, namely, a point between *my* eyes and directions extending from *my* eyes. In this way, the self is baked

---

[4] Burge (2019, 2022) also uses egocentric reference frames to account for the role of the self in perceptual representation.

into the accuracy conditions of not just this representation used with the cyclopean reference frame, but any perceptual representation used with an egocentric reference frame. We thus see how the self's implicit role in perception is secured by the fact that the use of egocentric reference frames is architecturally encoded in our perceptual systems.

While understanding the use of egocentric reference frames in perceptual processing sheds light on how the self is implicit in perceptual representation, *Implicitness* claims that the self is *always* implicit in perceptual representation. In the next section, I challenge this view by showing that some perceptual representations include a constituent representation of self.

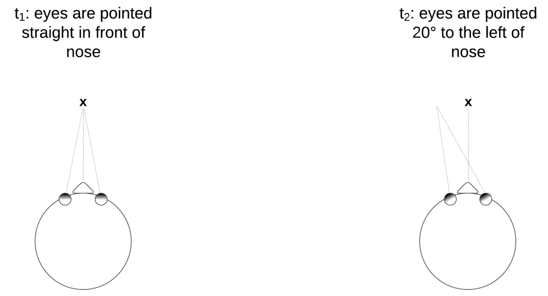
2. Counterexamples to *Implicitness*

In this section, I argue that *Implicitness* admits of two kinds of counterexamples. In both, the best explanation of the observed facts about perception is that one's complex perceptual representation includes a constituent representation of self.

2.1. Counterexample one: Gaze shifts

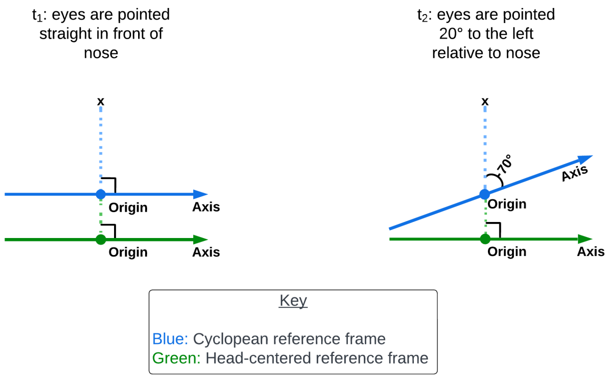
The first counterexample to *Implicitness* involves the routine perceptual experience of shifting your gaze. Let us try a short exercise, shown in Figure 1. Look straight in front of your nose, such that object x is in the center of your visual field. Call this t_1. Holding the position of your head 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. Further notice that if you try to grab x at t_2, you will reach in front of your head and not in front of your eyes. These facts give us reason to believe that you represent x's 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 keeping track of the position of your eyes relative to your head. But as I will argue, the best way of fleshing out the details of this account is incompatible with *Implicitness*.

Figure 1: 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. The origin of the *head-centered reference frame* corresponds to the center of one's head and the frame's axes correspond to directions extending from this point. As shown in Figure 2, at t_1, x is represented 90° from the cyclopean reference 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 reference 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 reference 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 in the cyclopean reference frame. Thus, at t_2, x is still represented 90° from the left-right axis of one's head-centered reference frame.

Figure 2: A representational account of gaze shifts



What this explanation of gaze shifts highlights is that x's position in one's head-centered reference frame is a function of x's position in one's cyclopean reference frame and the position of the cyclopean frame in one's head-centered frame. The problem for *Implicitness* is that plausibly, the visual system performs this calculation by using representations of each of the relevant arguments. In other words, the visual system computes x's position in one's head-centered reference frame using a representation of x's position in the architecturally encoded cyclopean reference frame and a representation of the cyclopean frame's position in the architecturally encoded head-centered frame. It is use of the latter representation that violates *Implicitness*. This is because a representation of the cyclopean reference frame is a kind of representation of one's eyes and so, a kind of representation of self.

One might reject that the cyclopean reference frame is represented during gaze shifts and offer the following account instead. Perhaps the function used to compute the position of object x relative to the head-centered reference 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° counterclockwise, the location of object o relative to the head-*

centered reference frame = the location of o in the cyclopean reference frame + x°. If this kind of rule is encoded, then no representation of the position of the eyes is necessary. How objects are specified relative to the head-centered reference frame is guaranteed by how the visual system functions. This counter-proposal not only respects *Implicitness*, but also suggests that a significant amount of information used in perceptual processing remains implicit.

Given this counter-proposal, it seems that theoretical considerations fail to adjudicate the question of whether a representation of eye position is used during gaze shifts. On the one hand, the variability of the eyes' position relative to the head suggests that it is represented. On the other, this variability can be accounted for in the visual system's functional architecture, bypassing the need for a representation. While it may seem that we are at an impasse, empirical data tips the scales. I turn to discuss two studies of visual perception, which considered together, lend support to positing a representation of eye position.

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 which appeared sensitive to eye position (pp. 640-641). For each of these neurons, the strength of the neural signal seemed to correlate with a particular direction of the gaze relative to the head. For example, one neuron exhibited baseline activity for all positions below the center of the gaze, increasing activity for positions above the center, and maximum activity for 15° from the vertical axis and 0° from the horizontal axis. This neuron thus seemed tuned to a gaze position of 15° from the vertical axis and 0° from the horizontal axis (Wang et al., 2007, p. 641). 70% of the neurons also showed a phasic response: when the monkey made a saccade to the neuron's preferred gaze position, the signal would initially spike, before exhibiting a constant firing rate for the duration of time at which the monkey held a fixed gaze (Wang et al., 2007, p. 642). The correlation between neural responses and gaze positions led the researchers to conclude that there is a 'representation' of eye position in the somatosensory cortex of the monkey.[5]

For good reason, Wang et al. (2007) did not discuss the possibility that the correlation between neural responses and gaze positions might be evidence that information about eye position is architecturally encoded, rather than represented. Functional architecture encodes algorithms or rules which describe how a system transitions from one representational state to another (Pylyshyn, 1984). Given this, we should not expect to look *within* a neural state of the brain and find a neuron that instantiates a functional rule. We should need to look at the transitions *between* neural states to infer a functional rule.

Wang et al. (2007) attempted to determine the origins of the eye position representation. One possibility was that it 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. (2007) 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 gaze 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.

---

[5] In most 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).

concluded that the gaze position signal was of proprioceptive origins, and that this signal was cut off when the eye was anesthetized.

While it provides evidence of a proprioceptive representation of eye position, Wang et al. (2007)'s study does not give us reason to believe that 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 & Miall (2008)'s results indicate just this.

In Balslev & Miall (2008)'s study, human subjects were tasked with performing a straight ahead test both before and immediately following rTMS (repetitive transcranial magnetic stimulation) over the left primary somatosensory cortex. In the straight ahead test, subjects 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). While performance was highly accurate in the control condition, performance was inaccurate following fifteen minutes of rTMS (Balslev & Miall, 2008, p. 8970). rTMS affected subjects' visual perceptions, shifting the visual scene to the left of their nose.[6] Balslev & Miall (2008) wrote 'we interpret this shift as an error in perceived eye position' (p. 8970).[7] 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.

One shortcoming of Balslev & Miall (2008)'s study is that rTMS occurs over a relatively large area, and thus, it could be that the observed changes to visual perception resulted from more generalized changes to connections both within and without the somatosensory cortex. However, as they themselves 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 and that it is used to locate objects relative to the head.[8] If this is correct, it seems reasonable to conclude that this representation is used during gaze shifts.[9]

To recap, what we wanted to explain is how during gaze shifts, one is able to represent object positions relative to the head. I proposed that representing an object's position relative to the head-centered reference frame required representing the cyclopean reference frame's position – i.e. the eyes' position – relative to the head-centered frame. I found support for this idea by looking at two empirical studies of visual perception, which together confirm that a representation of eye position is used to locate objects relative to the head. The use of a representation of eye position constitutes use of a representation of self. So, I conclude that gaze shifts involve use of a representation of self and as such, are a counterexample to *Implicitness*.

---

[6] 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 & Miall, 2008, p. 8970).

[7] rTMS over the left motor cortex had no effect on accuracy in the straight ahead test, confirming the source of the error as proprioceptive rather than motor (Balslev & Miall, 2008, p. 8970).

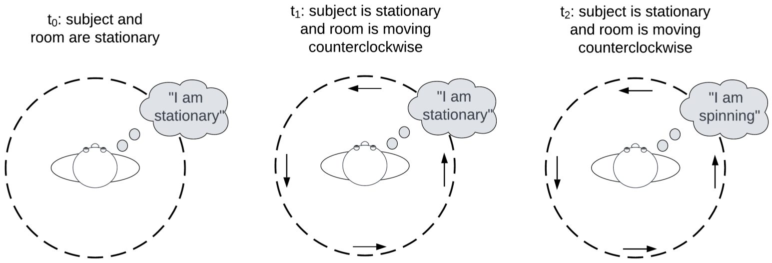
[8] The view that certain correlations between world states and internal states are sufficient to postulate mental representations is articulated and defended in Gallistel (1998) and Shea (2018, chap. 4).

[9] One might point out that this representation of eye position is proprioceptive and that proponents of *Implicitness* do not deny that we have proprioceptive representations. But the use of proprioceptive representations *per se* is not what is at issue. What is at issue is that *Implicitness* seems to deny or overlook that proprioceptive representations can be constituents of complex perceptual representations.

2.2. Counterexample two: Vection

The second counterexample to *Implicitness* that I will look at involves 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 Figure 3, runs as follows (see Mach, 1875/2001). A subject is seated inside a drum painted with black and white vertical 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 2-20 seconds after the drum begins to spin (Riecke & 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 stationary drum. Call this t_2. I will argue that the best representational explanation of the shift in the subject's experience between t_1 and t_2 violates *Implicitness*.

Figure 3: Vection

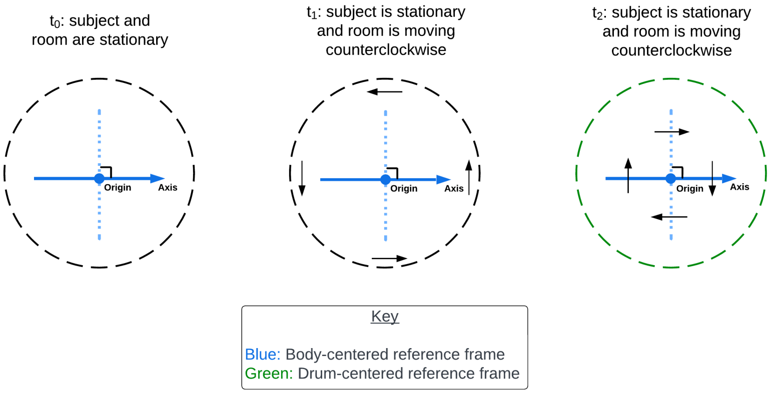


Vection is often understood as involving a switch in one's *rest frame* – a reference frame that a subject takes to be stationary (Prothero & Parker, 2003; Riecke, 2011; Seno et al., 2009). The idea is that one's rest frame at t_2 differs from one's rest frame at t_0 and t_1. Let us see how this works. 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. The origin of the body-centered frame corresponds to the center of the subject's body and its axes correspond to directions extending from this point. As shown in Figure 4, at t_0, the drum is represented as stationary in the body-centered reference frame and at t_1, the drum is represented as spinning counterclockwise in the body-centered 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. All spatial relations which were previously specified relative to the subject now are specified relative to the drum.

The problem for *Implicitness* is that the drum-centered reference frame is not used on its own: the subject

Figure 4: A representational account of vection



represents herself as spinning in this frame.[10] While I will consider shortly how we might spell out the idea that the subject represents *herself* as spinning, all we need to recognize for now is that the representation used is likely a kind of representation of self. Thus, vection seems to constitute a counterexample to *Implicitness*.

But an alternative explanation of vection poses no threat to *Implicitness*. Rather than positing the use of a representation of self, perhaps we can construe information about the subject spinning relative to the drum-centered reference frame as architecturally encoded in the subject's visual system. Recall that in our discussion of gaze shifts, we considered a parallel suggestion, namely, that the visual system encodes a rule to compute object positions in the head-centered reference frame. Perhaps a similar rule concerning the spatial relations between oneself and the drum is used during vection.

While it respects *Implicitness*, this counter-proposal is not viable for two reasons. First, to avoid use of a representation of self, presumably the rule must compute – i.e. output – the position of the drum relative to the architecturally encoded body-centered reference frame. But a representation of the drum in the body-centered frame is what we posited is used at t_1. Thus, this suggestion seems to collapse any representational difference between t_1 and t_2. Moreover, even if we assume that the complex representation used at t_2 somehow differs from that used at t_1, the spatial relations between oneself and the drum are too variable to be fit for architectural encoding. What made the analogous counter-proposal in the case of gaze shifts plausible – although it was ultimately rejected given the neural evidence – was that the movements of one's eyes are constrained, making it relatively straightforward to encode their varying position within the head. By contrast, the spatial relations between oneself and one's environment are unconstrained. 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. Computational cost is another reason to reject that information about one's position relative to the drum-centered reference frame is architecturally encoded.

---

[10] 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.

We are left with the intuitive explanation that at t_2, the subject represents herself as spinning in the drum-centered reference frame. But what kind of representation of self is used? The most parsimonious posit is that the subject's position relative to the drum-centered reference frame is given by a representation of an egocentric reference frame. Representing an egocentric reference frame relative to the drum-centered reference frame is sufficient to specify the subject's location relative to the drum, without additionally specifying the size and shape of the subject's body. As shown in Figure 4, I take it that the relevant egocentric reference frame is the subject's body-centered reference frame and that at t_2, it is represented as spinning in the drum-centered frame. Thus, an oriented point – in this case, the center of the subject's body – is represented as spinning relative to the drum.

To sum up, I have argued that vection involves the use of a representation of self. Specifically, I have suggested that when the illusory experience of self-motion begins, one represents one's body-centered reference frame as spinning in a drum-centered reference frame. Because a representation of the body-centered reference frame is a representation of self, vection constitutes a second counterexample to *Implicitness*.

Interestingly, this analysis of vection parallels our analysis of gaze shifts. Recall that in gaze shifts, we took one to represent one's cyclopean reference frame in a head-centered reference frame, and we understood a representation of one's cyclopean frame to be a representation of self. What we thus find is that in both gaze shifts and vection, information about the self is explicit when one egocentric reference frame is represented in another reference frame. In the next section, I use this idea to propose a new account of the representation of self in perception.

3. The *Nested Frames View*

Our analyses of gaze shifts and vection reveal that *Implicitness* admits of counterexamples: perceptual representations sometimes include constituent representations of self. This gives us reason to reject *Implicitness*. But in rejecting *Implicitness*, we are left without a systematic account of when and how constituent representations of self are used in perception. In this section, I give such an account. I propose the *Nested Frames View*, which describes when and how egocentric reference frames are represented in perception, and I discuss how this view elucidates the use of representations of self in perception.

The *Nested Frames View* holds that there are systematic relations between the reference frames used in perception and describes those relations thus:

> ***Nested Frames View:*** If multiple reference frames are used by a creature in perception, they exhibit an iterative nested structure.[11]

Loosely speaking, the *Nested Frames View* takes reference frames to be like Russian dolls. Just as one Russian doll is nested within another, one reference frame is nested within another. To say that reference frame x is *nested within* reference frame y is to say that x is represented in y. This means that there is a representation of x at a given location, which in conjunction with the relevant functional architecture, represents x's location

---

[11] Hybrid reference frames (Carrozzo & Lacquaniti, 1994), idiosyncratic reference frames (Chang & Snyder, 2010), and somatosensory maps (Medina & Coslett, 2010) are not intended to fall within the scope of the *Nested Frames View*.

in or relative to *y*. In this way, when *x* is nested within *y*, *x* is represented and *y* is architecturally encoded. The nesting relation between reference frames *x* and *y* is shown in the bottom row of Figure 5. According to *Nested Frames View*, this nesting relation iterates, similarly to the nesting relation among Russian dolls. When the nesting relation iterates, reference frame *y* is itself nested within a third reference frame *z*, meaning that *y* is represented in *z*. The nesting relation between reference frames *y* and *z* is shown in the top row of Figure 5. The number of iterations of the nesting relation depends on the number of reference frames used by a creature in perception. Were the relation further iterated, pairs of reference frames – one nested within the other – would be stacked where ascending dots are depicted in Figure 5.

Figure 5: Reference frames in an iterative nested structure

Iterating the nesting relation entails that some reference frames are both encoded in functional architecture and represented. Consider the use of reference frame *y* in the nested structure depicted. As explained, when *x* is represented, it is represented relative to *y*, which is architecturally encoded. But given the nested structure, *y* itself is represented relative to *z*. That *y* is both architecturally encoded and represented reflects that it is used in two ways: it is used as the reference frame relative to which *x*'s position is specified and it is used to specify its own position relative to *z*. We saw this in gaze shifts and vection. In gaze shifts, the cyclopean reference frame is architecturally encoded when used as the reference frame relative to which object positions are specified, but represented when used to specify its own position relative to the head-centered frame. Similarly, in vection, the body-centered reference frame is architecturally encoded when used as the reference frame relative to which the position of the drum is specified, but represented when used to specify its own position relative to the drum-centered frame.

The dual use of egocentric reference frames tracks the self's roles as subject and object of perception. An egocentric reference frame captures the self's role as the subject of perception when architecturally encoded. For example, when architecturally encoded, the cyclopean reference frame is used to specify the locations of objects relative to one's eyes, which are perceiving the world. By contrast, an egocentric reference frame captures the self's role as an object of perception when represented. When the cyclopean reference frame is represented, it is used to represent the position of the eyes – now the objects of proprioception – relative to the head. Of course, even when they capture the self's role as an object of proprioception, representations of egocentric reference frames are used as constituents of complex perceptual representations. In this way, as a whole, the nested structure of reference frames is marshalled in support of the self's role as the subject of perception.

What the nested structure shows is that information about the self flips between being implicit and explicit. Each architecturally encoded egocentric reference frame within a nested structure constitutes implicit information about the self. And, each represented egocentric reference frame – each frame nested within another

– constitutes explicit information about the self.

How are reference frames arranged within a nested structure? This is determined by the degree of processing perceptual inputs undergo before they are represented in a particular reference frame. If inputs undergo very little processing before being represented in a particular reference frame, the frame will fall towards the bottom of the structure (like reference frame *x* in Figure 5). Conversely, if inputs undergo significant processing before being represented in a particular reference frame, the frame will fall towards the top of the structure (like reference frame *z* in Figure 5). Perceptual inputs thus will flow from the bottom of the structure up.

Given this, we can sketch what the nested structure used in human perception might look like. For simplicity, let us assume that only four reference frames are used in human perception: the cyclopean reference frame, the head-centered reference frame, the body-centered reference frame, and an allocentric reference frame (which is a reference frame whose origin does not correspond to a point on the subject's body). If how these are nested is determined by degrees of perceptual processing, 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. 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 so on. Of course, parts of this nested structure – the nesting of the cyclopean frame in the head-centered frame, and the nesting of the body-centered frame in an allocentric frame – were revealed in our analyses of gaze shifts and vection.

This structure gives us a sketch of when and how representations of self are used in human perception. The cyclopean reference frame's nesting within the head-centered frame marks a representation of 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 a representation of 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 a representation of one's body used to specify its position relative to the environment.

In addition to tracking degrees of perceptual processing, this nested structure used in human perception mirrors the structure of bodily joints.[12] To see this, let us compare the relation between the cyclopean and head-centered reference frames to the relation between the eyes and the head. The nested structure is such that the head-centered reference frame always functions as the anchor relative to which variable positions of the cyclopean frame are represented. In other words, the cyclopean reference frame is represented at different positions in the architecturally encoded head-centered frame, but given the nested structure, the converse is not possible. Similarly, the eye sockets, which are a type of joint, enable the eyes to move when the absolute position of the head is fixed. But the converse is not possible: we cannot move our heads while keeping the absolute position of our eyes fixed. Like the head-centered reference frame's relation to the cyclopean reference frame, the head functions as the anchor relative to which the eyes can move. Similar parallels can be drawn between other nested frames and human joints.

While much more could be said about the details of the *Nested Frames View* and how it applies to different types of creatures, the discussion offered suffices for our purposes.[13] By describing the nested structure

---

[12] See Bermúdez (2017) on the B-locations of bodily events for a similar idea.

[13] For example, *which* and *how many* reference frames comprise a given nested structure will depend on things such as the creature's

exhibited by the reference frames used in perception, the *Nested Frames View* reveals information about the self to flip between being implicit and explicit. What we have seen is that information about the self remains implicit when an egocentric reference frame is encoded in the functional architecture of one's perceptual systems, while representations of self are used when one egocentric reference frame is represented in – i.e. nested within – another.

4. The hierarchy of representations of self

Having explicated the *Nested Frames View* and discussed how it provides an account of the use of representations of self in perception, I will argue that the structure of reference frames the view describes is a hierarchy. This is true for two reasons. First, the nested frames structure tracks increasing degrees of a particular kind of agency. Second, information from reference frames at the top of the structure contributes more to one's self-locating judgments than does information from frames at the bottom. I turn to develop each of these ideas in turn.

Before explaining the relation between the nested frames structure and agency, I want to sketch the kind of agency that I have in mind. This kind of agency – representational agency – is intended to apply to creatures that lack conceptual capacities (Burge, 2010, ch. 8). I will say that a creature has a capacity for *representational agency* if it has the capacity to perform representational actions. I further stipulate that the capacity to perform representational actions requires the capacity to form action representations of physical events and successfully act on those representations. We can think of an action representation as the non-conceptual analogue of an intention: it has the same function as an intention, but is specified without the use of conceptual representations. The qualification that the action representation be of a physical event is meant to exclude from our notion of representational agency cases of mental agency.

With this notion of representational agency in hand, I contend that we should expect creatures that lack a capacity for representational agency not to use any nested reference frames. Allow me to explain. I conjecture that a creature will use more than one reference frame in perception only if it has the capacity for movement. This is because multiple reference frames are useful only if they can 'come apart' – only if the spatial relations between the points and directions to which they correspond can vary. For example, the reason that both the cyclopean and head-centered reference frames are useful in human perception is because humans can move their eyes, and when they do, the directions corresponding to the axes of the cyclopean reference frame vary independently of the directions corresponding to the axes of the head-centered frame. The relation between the use of reference frames and the capacity for movement suggests that creatures that lack the capacity for movement will not use multiple reference frames, and so, will not make use of a nested frames structure. This establishes the bottom of our hierarchy: no representational agency, no nested frames structure.

Moving up the hierarchy, we should expect increasingly heavily nested structures – those with more reference frames – to track increasing degrees of representational agency. To understand why, we must recognize two things. First, creatures with more heavily nested structures generally have greater degrees of freedom

---

perceptual faculties, body shape, and possibilities for movement.

than creatures with less heavily nested structures. In other words, creatures with more heavily nested structures can move in more kinds of ways. This follows from the relation between the use of reference frames and the capacity for movement, discussed above. Because more reference frames are useful only when they might come apart in the sense described, we should expect nested structures to track bodily joints (as we saw for humans in §3). Thus, we should expect creatures with more heavily nested structures to have more joints and so, to have greater degrees of freedom in their movements.

Second, the use of increasingly heavily nested structures marks increasing amounts of proprioceptive information that is available to guide one's movements. The use of the head-centered reference frame marks the capacity to represent proprioceptive information about the movements of one's eyes; the use of the body-centered frame marks the capacity to represent proprioceptive information about the movements of one's head; and so on. In this way, each reference frame that is part of a nested structure comes with its own bit of proprioceptive information. Such information is critical to the successful performance of one's actions because it allows one to perceive one's actions while in progress and adjust them if and as necessary. So, we should expect creatures that use more heavily nested structures to act on their action representations successfully given the proprioceptive information available to them.

We now can tie these two strings together. Given their greater degrees of freedom, creatures with more heavily nested structures will be able to perform more kinds of representational actions, and given the proprioceptive information available to them, they will be able to perform these actions successfully. Thus, we should expect creatures that use more heavily nested structures to exhibit greater degrees of representational agency than creatures that use less heavily nested structures. This constitutes the first dimension in which the nested frames structure is hierarchical.

The second dimension in which the nested frames structure is hierarchical is that reference frames at the top of the structure are more likely than those at the bottom to be associated with one's conceptual representation of self (in creatures with conceptual capacities). Specifically, I conjecture that information from frames at the top is more likely to inform conceptual judgments about self-location.

Borrowing an example from Peacocke (1992), let us compare different ways in which you might look at Buckingham Palace. First, your eyes might be rotated relative to your head, such that the palace falls in the middle of your cyclopean reference frame, but off to left in your head-centered frame. My intuitions are that if prompted to judge where the palace is relative to you, you will say it is to your left. If so, your judgment will be based on the palace's location in your head-centered reference frame, rather than its location in your cyclopean frame. Of the two reference frames, it is the one higher in the nested structure that informs your judgment. The same seems true when we compare the use of your head- and body-centered reference frames. This time, imagine that your head is rotated relative to your body, such that the palace falls in the middle of your head-centered reference frame, but off to the left in your body-centered reference frame. Again, you seem more likely to judge that the palace is to your left and again, your judgment is based on the reference frame higher in the nested structure: it is based on the palace's location in your body-centered reference frame rather than in your head-centered frame.

There is some empirical confirmation of the hypothesis that one's body-centered reference frame more heavily contributes to one's self-location judgments than one's head-centered reference frame. Alsmith et al. (2017) used what they call a 'Misalignment Paradigm' to evaluate the relative contributions of the two reference frames. Over a series of 750 trials, subjects were tasked with judging whether a ball was to the left or right of an avatar viewed from a bird's-eye perspective. Because the avatar's head and torso were misaligned, Alsmith et al. were able to determine how the avatar's head- and body-centered reference frames contributed to subjects' judgments. While they found both reference frames to contribute, Alsmith et al. (2017)'s analysis revealed that 'the torso's [i.e. body's] overall contribution to judgments was substantially stronger than that of the head' (p. 110). This offers initial support for my hypothesis that one's self-locating judgments are more heavily influenced by information in reference frames towards the top of the nested structure.

To wrap up, the structure of reference frames posited by the *Nested Frames View* is hierarchical in two respects. First, it corresponds to increasing degrees of representational agency, such that creatures with more heavily nested structures exhibit greater degrees of representational agency than creatures with less heavily nested structures. Second, in creatures with conceptual capacities, reference frames at the top of a nested structure contribute more heavily to one's self-locating judgments than those at the bottom. Because the use of a nested frames structure marks the use of a series of representations of self, we can conclude that these representations of self form a hierarchy. The bigger the hierarchy, the greater one's representational agency. And the closer to the top of the hierarchy, the more a perceptual representation of self impacts the use of the conceptual representation of self.

5. Conclusion

I have offered a new account of how information about the self is used in perceptual representation. Specifically, I have proposed that such information flips between being implicit and explicit. Information about the self is implicit when an egocentric reference frame is encoded in the functional architecture of one's perceptual systems and explicit when one egocentric reference frame is represented relative to another. What this picture reveals is that many representations of self are used in perception. As demonstrated, this fits with empirically-grounded analyses of gaze shifts and vection, in contrast to the view that information about the self is always implicit in perception.

I have argued that the representations of self used in perception form a hierarchy that tracks increasing degrees of agency. I also have suggested that representations at the top of this hierarchy affect one's self-locating judgments more than those at the bottom. While more research is required to understand how this works, the tantalizing possibility is that the hierarchy of representations of self extends beyond perception and into thought.

References

Alsmith, A. J., Ferrè, E. R., & Longo, M. R. (2017). Dissociating contributions of head and torso to spatial reference frames: The misalignment paradigm. *Consciousness and Cognition*, *53*, 105–114.

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

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

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 Cambridge.

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

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

Burge, T. (2010). *Origins of objectivity*. Oxford University Press.

Burge, T. (2019). Psychological content and egocentric indexes. In A. Pautz, & D. Stoljar (Eds.) *Blockheads!: Essays on Ned Block's philosophy of mind and consciousness*, (pp. 41–69). MIT Press.

Burge, T. (2022). *Perception: First form of mind*. Oxford University Press.

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

Carrozzo, M., & Lacquaniti, F. (1994). A hybrid frame of reference for visuo-manual coordination. *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience*.

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

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

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, *22*(1), 319–349.

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

Fogassi, L., Gallese, V., Di Pellegrino, G., Fadiga, L., Gentilucci, M., Luppino, G., Matelli, M., Pedotti, A., & Rizzolatti, G. (1992). Space coding by premotor cortex. *Experimental Brain Research*, *89*, 686–690.

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

Graziano, M. S., Yap, G. S., & Gross, C. G. (1994). Coding of visual space by premotor neurons. *Science*, *266*(5187), 1054–1057.

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

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.

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

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).

Medina, J., & Coslett, H. B. (2010). From maps to form to space: Touch and the body schema. *Neuropsychologia*, *48*(3), 645–654.

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

Morgan, D., & Salje, L. (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.

Ono, H., & Barbeito, R. (1982). The cyclopean eye vs. the sighting-dominant eye as the center of visual direction. *Perception & Psychophysics*, *32*(3), 201–210.

Palmisano, S., Allison, R. S., Schira, M. M., & Barry, R. J. (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*. The MIT Press.

Peacocke, C. (2014). *The mirror of the world: Subjects, consciousness, and self-consciousness*. Oxford University Press.

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

Perry, J., & Blackburn, S. (1986). Thought without representation. *Proceedings of the Aristotelian Society, Supplementary volumes*, *60*, 137–166.

Prothero, J. D., & Parker, D. E. (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*. The 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., & Schulte-Pelkum, J. (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.

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

Seno, T., Ito, H., & Sunaga, S. (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.

Toribio, J. (2008). State versus content: The unfair trial of perceptual nonconceptualism. *Erkenntnis*, *69*, 351–361.

Wang, X., Zhang, M., Cohen, I. S., & Goldberg, M. E. (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*, vol. 34. Blackwell Oxford.

Zipser, D., & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature*, *331*(6158), 679–684.