UNIVERSITY OF CALIFORNIA

Los Angeles

Representations of the Perceiving Self

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Philosophy

by

Catherine Pettibone Hochman

2025

ABSTRACT OF THE DISSERTATION

Representations of the Perceiving Self

by

Catherine Pettibone Hochman
Doctor of Philosophy in Philosophy
University of California, Los Angeles, 2025
Professor Gabriel Jae Greenberg, Chair

Imagine you're apple picking and an apple on the branch in front of you catches your eye. Your visual system forms a representation of the apple, including not only its various properties, such as its color, shape, and texture, but also its location. Its location is specified egocentrically: you represent the apple at some distance and direction relative to yourself. This suggests that *you* are represented in your visual representation. But that seems strange. Intuitively, you visually represent the objects – not the subject – of perception. What's more, it is not clear what it would mean for you to be represented in your visual representation. Does your visual representation, like the thought *'An apple is in front of me'*, include the concept of self?

This case raises two questions about the relation between perception and the self. One asks *whether* the self is represented in perception. The other concerns *how* the self is represented in perception. This dissertation addresses both.

I answer the first question in the affirmative. I argue that perception involves the use of first-personal contents that are represented in perceptual systems that map complex spatial relations. This suggests that nearly all perceiving creatures self-represent in perception and challenges the

traditional view according to which self-representation is confined to language and thought.

However, how the self is represented in perception differs from its representation in language and thought. Perception involves the use of egocentric reference frames – frameworks for specifying spatial relations relative to the perceiver – that, I argue, carry first-personal contents. I draw on work in neuroscience and perceptual psychology to offer a representational trajectory that traces how these reference frames, and thus their contents, move from being encoded in the functional architecture of our perceptual systems to explicitly represented. Each step along this path moves closer to conceptual self-representation, suggesting that our capacities for self-representation in language and thought may have their origins in perception.

This dissertation of Catherine Pettibone Hochman is approved.

Joshua David Armstrong

Katrina Elliott

Pamela Hieronymi

Jenann Ismael

Gabriel Jae Greenberg, Committee Chair

University of California, Los Angeles

2025

Contents

List of Figures

Acknowledgments

I am deeply grateful to Gabriel Greenberg, without whom this dissertation would not have been possible. We began working together at the end of my second year, when he generously agreed to advise me on an underdeveloped idea at the intersection of philosophy of mind and cognitive science – even though I had almost no background in either. From the start, Gabe supported my philosophical interests and showed patience and commitment in helping me develop them. Over the many years since, he has worn all the hats of an exceptional mentor – teacher, taskmaster, cheerleader, and advocate – and I am grateful to him for taking on each role as needed. I am also indebted to Gabe and the rest of my committee for encouraging me to bet on myself in my professional choices. My path would be quite different had he not placed his faith in me and shown me that I could do the same.

I've been incredibly lucky to have an all-star committee. Josh Armstrong helped me consider how my project fit into the broader philosophical landscape whenever I started to lose sight of the forest for the trees. Of equal importance were his kind reminders that philosophy is one element of a broader life.

I have learned many lessons from Pamela Hieronymi, perhaps chief among them that becoming an academic philosopher is a lifelong pursuit that involves developing a set of skills over time. Her detailed feedback always hit the mark and inevitably resurfaced whenever I didn't address it fully.

Jenann Ismael, who kindly agreed to join my committee, has been an excellent interlocutor and frequently pointed me to areas in the literature deeply relevant to my research. She has also provided sage career advice, especially when I faced difficult decisions.

Last but not least, Katrina Elliott's mix of enthusiasm, humor, and clarity consistently opened the creative floodgates and led to some of the most intellectually stimulating moments of my graduate career. Beyond her role on my committee, Katie is also my best model for how to teach; I aspire to be like her in the classroom.

Asking these people to serve on my committee was the best decision I made during graduate school. I cannot thank them enough for their support, encouragement, kindness, and mentorship.

Many others have shaped my philosophical journey. William Everdell and Benjamin Rutter were the first to point me towards philosophy in high school. I would not have pursued graduate school without the guidance of my undergraduate mentors at Princeton: Thomas Kelly, Benjamin Morison, and Gideon Rosen. Benjamin Morison, who advised my junior paper and senior thesis, instilled in me a lasting love of philosophy (and Plato!). Tyler Burge, John Carriero, Sam Cumming, Barbara Herman, David Kaplan, Calvin Normore, Michael Rescorla, David Schaberg and Sean Walsh provided invaluable support throughout graduate school. Calvin Normore and David Schaberg, in particular, together helped foster my interest in Chinese philosophy; David patiently taught me the fundamentals of Classical Chinese, an experience I will always treasure. Allison Aitken, Ned Block, David Chalmers, Eric Mandelbaum, and Matthias Michel welcomed me to their classes and workshops during my stay in New York, which made that time especially fruitful.

My work on this dissertation benefited greatly from questions and feedback from audiences at the APA Central (February 2024), the International Society for the Philosophy of the Sciences of the Mind (November 2023), the CUNY TEEM Lab (November 2023), the Salzburg Conference for Young Analytic Philosophy (September 2023), the European Society for Philosophy and Psychology (August 2023), the Society for Philosophy and Psychology (June 2023), APA Pacific (April 2022), and the Dubrovnik Conference on Cognitive Science (2021). I am especially grateful to José Luis Bermúdez and Jonna Vance for their incisive feedback at two of those conferences.

Special thanks to the UCLA philosophy department's administrators, especially Doug Myers, David Pastor, and Isamara Ramirez, who made many aspects of my graduate experience run smoothly – from deadlines and applications to funding.

I have learned immensely from conversations with my peers, including Tristen Cardwell, Christian De Leon, Mona Fazeli, Gabbrielle Johnson, Kevin Lande, Jürgen Lipps, Alonso Molina, Torsten Odland, David Pederson, Alexi Patsaouras, Samuel Pensler, Jiarui Qu, Nefeli Ralli, Daniel Ranweiler, Kyle Scott, and Drew Stewart. I am also grateful to the many others at UCLA who helped me work through ideas at the Albritton Society, the Ethics Workshop, the Mind and Language Workshop, and the Proposition Series. My special thanks go to my close friends: Zachary Biondi and Paul Talma, for providing some of my most engaging and wide-ranging philosophical discussions; Alexandra Pelaez, for being my adventure buddy; Ekin Zeytinoglu, for being a wonderful work companion; and Colleen Hanson for being my rock through the final years of graduate school.

My deepest gratitude goes to my family. To Caroline, for being my greatest cheerleader and for always finding a way to make me laugh; to my parents, for everything; and to Will, for his constant support and love throughout this big undertaking. Here's to our next big adventure!

Curriculum Vitae

Employment

2025–2026	Postdoctoral Researcher, Université Grenoble Alpes

Education

2021	C.Phil. in Philosophy, University of California, Los Angeles
2018	M.A. in Philosophy, University of California, Los Angeles
2016	A.B. in Philosophy, *cum laude*, Princeton University

Publications

"The Hierarchy of Selves in Perception." *Review of Philosophy and Psychology*, forthcoming.

Presentations

2024	"The Hierarchy of Selves in Perception," APA Central Division, February 21–24
2023	"The Hierarchy of Selves in Perception," International Society for the Philosophy of the Sciences of the Mind, November 24–December 2
2023	"The Hierarchy of Selves in Perception," Salzburg Conference for Young Analytic Philosophy, September 6–8
2023	"The Hierarchy of Selves in Perception," European Society for Philosophy and Psychology (ESPP), August 28–31
2023	"The Hierarchy of Selves in Perception," Society for Philosophy and Psychology (SPP), Poster, June 21–23
2022	"Spatial Representations and the Self," APA Pacific Division, April 13-16
2021	"Indexical Information in Perceptual Representation," Dubrovnik Conference on Cognitive Science, May 18–21

Chapter 0

Introduction

Imagine that I'm playing fetch with my dog in the backyard. I throw the tennis ball, and he chases after it, tracking its arc through the air. While we're both engaged in the same activity, our experiences are presumably quite different. First, I am self-conscious: I'm aware not only of the ball's flight but also of my role as the one who threw the ball. I also possess self-knowledge: I know that I'm playing fetch with my dog, that I like playing with him, that I'm in my backyard, and so forth. I can exercise agency. I can weigh various considerations – whether to throw the ball again, how hard to throw it – and decide how to act. Finally, when I exercise this agency, I bear responsibility: if I were to chuck the ball at my neighbor walking by, I would be responsible for my action. Plausibly, none of these things is true of my dog: he is not self-conscious, nor does he have self-knowledge, agency, or moral responsibility.

What accounts for this difference? A natural explanation is that I can represent myself while my dog cannot. The idea is that while my mind carries information about my self, my dog's mind does not, and this difference is relevant, or even critical, to my being aware of myself, to my knowing about myself, to my choosing how I will act, and to my responsibility for my actions. To flesh this out further, imagine that the mind is a filing cabinet that contains different files about different aspects of the world. That my mind includes a file about my self and my dog's mind does not may help explain why I, for example, am self-conscious and he is not.

The reason to think that my dog cannot self-represent is that he presumably lacks capacities for language and conceptual thought. The traditional picture locates self-representation exclusively in these domains. In English, we represent ourselves using words like “I,” “me,” and possibly “mine.” In thought, we represent ourselves whenever we think about ourselves. When I think *‘I should throw the ball again,’* my thought involves a representation of my self. But if my dog does not have the same general capacities for language and thought, then he seems precluded from self-representing.

I argue against this traditional picture. The central thesis of this dissertation is that self-representation extends to perception. To motivate this idea, note that perceptual experiences are egocentric. When I look at the ball, for example, I locate it relative to my self. This suggests that my visual representation includes a representation of self. That is, in the same way that the utterance “Get ready, I’m going to throw the ball!” and the thought *‘I’m only throwing a few more’* involve representations that refer to the speaker and thinker, respectively, my perceptual state when I see the ball involves a representation that refers to the perceiver (i.e., my self). This reasoning equally applies to other perceiving creatures. Most perceiving creatures, according to my view, self-represent in perception.

If this is right, then the primary building block of self-consciousness, self-knowledge, agency, and moral responsibility may be both more primitive and more widespread than we typically suppose. This is because the representation of self found in perception may mark the origins of the representations of self used in thought and language. That is, our capacities to self-represent in language and thought might be bootstrapped from more basic self-representational capacities found in perception. This follows from the idea that perception is the most primitive type of mental representational system; it is where the “representational mind begins” (Burge, 2014, p. 396). Thus, by investigating the representation of self in perception, we may gain greater insight into its workings in thought and language.

This idea that the origins of the representation of self used in language and thought might lie in perception has garnered growing attention in the philosophical literature over the last few decades.[1] This dissertation contributes to this line of inquiry by offering what we might consider a "representational trajectory" of how representations of self are used in perception. By "representational trajectory," I mean an account that traces how information about the self moves from being encoded in the functional architecture of our perceptual systems to being explicitly represented.

I summarize the broad strokes of this trajectory here before offering a more detailed account in the chapter summaries below. To begin, first-personal contents are implicit in perceptual representations of the most basic spatial relations. What's more, such contents are architecturally encoded in perceptual systems; they are trapped within the walls of the mind. Once a creature begins to represent more complex spatial relations, first-personal contents become explicitly represented. These explicit representations form a hierarchy, where first-personal information that is implicit at one level becomes explicit at the next. Finally, a single representation of self – rather than a hierarchy – is used to represent the body that feels like one's own. In this way, perceptual systems move from using architecturally encoded first-personal contents, to using multiple hierarchically arranged representations of self, to using a single representation of the whole bodily self. Each step along this path inches closer to the conceptual representation of self found in thought.

In Chapter One, I ask whether perceptual representations carry first-personal contents by examining their role in action explanations. The dominant view in the literature holds that action explanations require that perceptual representations carry first-personal contents, while a more recent proposal suggests that they carry merely indexical locational contents that refer to the perceiver's location. I argue that both views capture something essential that the other neglects: first-personal contents explain how perception motivates action, while indexical locational contents explain perceptual-motor coordination. Since each view faces explanatory challenges when taken alone, I propose the Located Individual View, which holds that perceptual representations include both first-personal and indexical locational contents. I argue that this hybrid account can

[1] See, for example, Bermúdez (1998), Burge (2019), Hurley (1998), Musholt (2015), and Peacocke (2014).

best explain how perception both motivates and guides action.

I then propose that perceptual representations include both first-personal and indexical locational contents in virtue of perception's use of egocentric reference frames – frameworks for specifying spatial relations relative to the perceiver. I present a unified account of these frames, which are central posits in perceptual psychology, in Chapter Two. First, I use a Kaplanian framework to explicate how these frames might carry both first-personal contents and indexical locational contents. Second, I contend that when used in their standard role as anchors for specifying spatial relations, egocentric reference frames remain implicit – they contribute to the accuracy conditions of perceptual representations without being explicitly represented. Third, I suggest that egocentric reference frames are encoded in the functional architecture of perceptual systems. Taken together, these theses entail that first-personal contents are encoded in the architecture of our perceptual systems.

Are first-personal contents always encoded in the architecture of our perceptual systems? In Chapter Three, I argue that this 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 architecturally encoded. 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 first-personal contents become 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 marshal empirical evidence and theoretical considerations in support of this hierarchical picture and show how it unifies standard explanations of two perceptual phenomena: gaze shifts and the illusory experience of self-motion.

In Chapter Four, I examine the move from the use of many, hierarchically arranged, representations of self to the use of a single representation of self. Here, I turn my focus from exteroception to the sense of bodily ownership, the feeling that our bodies are our own. I first discuss an account of the sense of bodily ownership offered by Bermúdez (2017), according to which we represent the space of our bodies using a hierarchy of body part representations. While this account seems to explain the sense of local body ownership – the sense of ownership we have for our body parts, I argue that emerging neural and psychological evidence suggests it cannot explain the sense of global body ownership – the sense of ownership we have for our whole bodies. To account for the latter, I propose that we posit a single representation that is used to represent objects as belonging to one's self. I end by briefly comparing this representation to the conceptual representation of self used in thought.

Chapter 1

Locating the Self: Indexicals in Perception and Action

Our actions are guided by, and thus partially explained by, our perceptions. When my friend tosses me a frisbee, I am able to catch it in part because of perceptual information about its location. But how, exactly, is its location perceptually represented in a way that enables successful action?

In contemporary philosophy of perception, perceptual representations are often described as "self-locating."[1] That is, however I represent the frisbee's location, I also, in some sense, locate myself. According to certain articulations of this view, my perceptual contents take the form *frisbee approaching me* and so involve a *de se* or first-personal element.[2] An alternate view of perceptual contents seemingly replaces this *de se* element with a spatial indexical. The idea is that my perceptual contents instead take the form *frisbee approaching here*.[3] It is generally assumed

[1] See Alsmith (2017), Ávila (2014), Brewer (1992), Cassam (1997), Fernandez Velasco (2024), Schellenberg (2007, 2016), and Schwenkler (2014) for explicit discussions of the idea that perception is self-locating.

[2] That my perceptual contents take the form *frisbee approaching me* leaves substantial room for interpretation. One question is what exactly the *de se* element refers to. Another question is whether perceptual contents that include a *de se* element also include a spatial indexical. These complexities are discussed in greater depth in §1.3 and §1.4.

[3] Whether this view might be grouped with those that take perceptual representations to be self-locating is disputed. Mitchell (2021), for example, argues that perceptual contents involve only a spatial indexical and clearly contrasts this with the "self-locating" view according to which perceptual contents involve a *de se* element. By contrast, Schellenberg (2016, p. 339-341), who also defends the idea that perceptual contents can involve only a spatial indexical, thinks this is one way of unpacking the idea that perception is self-locating.

that the first view can explain how I manage to catch the frisbee (see Bermúdez, 1998). More recently, it has been argued that the second view can offer an even simpler explanation (Mitchell, 2021; Schellenberg, 2016).

In this chapter, I argue that both views of perceptual contents face serious challenges when used to explain action. The view that posits purely *de se* contents struggles to account for how perceptions can inform action; the view that posits purely indexical locational contents struggles to explain how perceptions can motivate action. To address both of these problems, I propose that perceptual representations include both *de se* contents and indexical locational contents. More specifically, I suggest that there is a single representational element in perceptual representations that simultaneously expresses *de se* and indexical locational contents.

I proceed as follows. In §1.1, I provide a more detailed account of both views of perceptual contents and discuss how each is thought to contribute to explanations of action. In §1.2, I begin by distinguishing two desiderata that I argue any representational-level explanation of action should meet. I then contend that the view that posits indexical locational contents struggles to satisfy the second desideratum. In §1.3, I discuss how the view that posits *de se* contents struggles to satisfy the first desideratum. Finally, in §1.4, I articulate the Located Individual View and argue that it satisfies both desiderata.

1.1 Two views of perceptual contents

Imagine that I'm playing frisbee with a friend. She tosses the disc and I watch it fly towards me. Looking at the frisbee, I visually represent not only its properties, such as its color, shape, and texture, but also its location relative to myself. But how exactly do I represent the frisbee's egocentric location? One possibility is that I represent the frisbee at some distance and direction relative to my self. If so, then I, like the frisbee, figure in my visual contents. This intuitive thought comports with the more general view, widely suggested in the literature, that the perceiver figures in the contents of her egocentric perceptual representations (Alsmith, 2017; Bermúdez, 1998; Brewer, 1992; Cassam, 1997; Hurley, 1998; Peacocke, 1999; Schwenkler, 2014).

That perceptual contents are first-personal in this way is often taken as an explanandum. The standard explanation is that perceptual contents are made first-personal in virtue of their connection to action (Alsmith, 2017; Bermúdez, 1998; Brewer, 1992; Cassam, 1997; Hurley, 1998). Brewer (1992), for example, articulates this idea by saying that "it is, at least in part, the way in which perception is taken up in the guidance and control of the flexible, world-behavioral responses of a single, persisting physical entity, which constitutes its egocentric, self-locating spatial content" (p. 28). According to this view, perceptual contents would not be first-personal if the link between perception and action were severed. They are first-personal precisely because they contribute to "a subject's capacity for basic purposive action in the world" (Brewer, 1992, p. 26).

Of course, this explanation is moot if perceptual contents aren't first-personal, and perhaps they aren't. What if I represent the frisbee at some distance and direction relative to my location? If this is the case, it is my location, rather than my self, that figures in my visual contents. This alternative possibility reflects the broader view that perceptual representations of egocentric spatial relations do not necessarily involve *de se* contents; instead, they can represent such relations using an indexical representation of a location, that which is in fact occupied by the perceiver (Mitchell, 2021; Peacocke, 2014, 2016; Schellenberg, 2016).[4]

Both views agree that perceptual contents, understood as something like Fregean senses or modes of presentation, are indexical. They differ in what they index. According to what I will call the *Individual Contents View*, perceptual representations have indexical contents that refer to the perceiver. One way to understand this is by analogy to first-person thought. Just as my thought '*A frisbee is approaching me*' involves indexical contents that refer to the thinker (i.e., me), my

[4] A related debate, which I take up in Chapters Two and Three, concerns whether first-personal perceptual contents are implicitly or explicitly represented. Campbell (1994, p. 119), Perry (1986), Musholt (2015, pp. 80-81), and Recanati (2012, pp. 185-190) articulate the view that first-personal perceptual contents are implicit; Schwenkler (2014) argues that such contents are sometimes explicit. Like Peacocke (2014, p. 32), I take this question concerning the implicit/explicit distinction to be orthogonal to the question of whether perceptual contents are first-personal. Peacocke (2014) frames the latter question in terms of a difference between what he calls "Degree 0" and "Degree 1." Creatures at Degree 0 are not capable of any kind of self-representation – they merely represent locations – while creatures at Degree 1 are capable of non-conceptual self-representation.

perceptual representation of the frisbee involves indexical contents that refer to the perceiver (i.e., me). By contrast, what I will call the *Locational Contents View* takes perceptual representations to have indexical contents that refer to a location. While the location referred to is that occupied by the perceiver, the perceiver does not appear in the contents or extension of her perceptual representations. According to this view, the indexical element of perceptual contents is analogous to the content expressed by the conceptual representation '*here*,' not the content expressed by the conceptual representation '*I*.'

The locus of debate between these two views of perceptual contents centers on the role that perceptual contents play in explanations of action. Those who take perceptual representations to have indexical individual contents argue that this view of contents is necessary to explain action. The rejoinder is that action can be explained by positing merely indexical locational perceptual contents instead. To frame this debate, I first discuss Perry (1979)'s famous case of the messy shopper, which is used to illustrate why *de se* beliefs seem necessary to explain action. I then show how this reasoning is carried over to perception and used to argue that indexical individual perceptual contents are necessary to explain action. Lastly, I explore how some reject this in favor of an allegedly simpler explanation of action that posits merely indexical locational perceptual contents.

Perry (1979) argues that *de se* beliefs are necessary to explain action by analyzing the following case. He imagines spotting a trail of sugar on the floor of a supermarket and forming the belief, *'The shopper with the torn sack is making a mess.'* After following the trail for some time, Perry realizes that *he himself* is the culprit with the torn sugar sack. He now believes *'I am making a mess'* and so rearranges the sack in his cart. Both his first-personal *de se* belief and third-personal *de dicto* belief are made true by the same fact – that Perry is making a mess. Nonetheless, Perry argues that with his *de se* belief he comes to believe something new, so the two beliefs must be different. Their difference is highlighted by the apparent fact that only the *de se* belief provides an adequate explanation of Perry's behavior. It is only by referencing Perry's *de se* belief that we can explain why he stops following the sugar trail and rearranges the sack in his cart.[5] Perry concludes that *de se* beliefs are essential to explanations of behavior.

We can adapt Perry's style of argument to make the case that perceptual representations with indexical individual contents are necessary to explain action. Imagine that I'm looking in a mirror and see a frisbee flying towards the person in the reflection. Initially, I don't recognize the person as myself. My visual representation has the rough content, '*A frisbee is approaching that person*.' Then, I suddenly recognize that the person in the mirror is myself. My visual content changes to *'A frisbee is approaching me'*, and I quickly turn to catch the disc. Both visual representations are made true by the same fact – that a frisbee is approaching me. So, what explains my change in behavior? Plausibly, a change in my visual contents. Only the second representation, which includes indexical individual contents, adequately explains why I turn to catch the frisbee. Like the *de se* belief *'I am making a mess'*, only this first-personal representation describes how I fit into my environment in the way necessary to inform my actions. This seems to suggest that perceptual representations with indexical individual contents are essential to explanations of behavior.

Bermúdez (1998) uses this kind of argument to explain the actions of non-linguistic creatures. He suggests that "in cases where the behavior of non-language-using creatures demands an intentional explanation, such explanations can draw on first-person perceptual contents just as explanations of the behavior of more conceptually sophisticated creatures can draw on first-person beliefs" (p. 118). In other words, we can run a Perry-style argument to explain the behavior of creatures that lack linguistic, and presumably, conceptual, capacities. To explain why a mouse scurries into a crevice upon spotting a predator, for example, we can posit that the mouse visually represents the predator at a location relative to its self using indexical individual contents. Our reason for positing indexical individual contents in this case is analogous to our reasons for positing such contents when I catch the frisbee and when Perry adjusts the sugar sack in his cart.

[5] Kaplan (1989) illustrates this point using a different example: one's actions will vary depending on whether one thinks *'His pants are on fire'* or *'My pants are on fire'*, even when both thoughts are made true by the same fact (p. 533).

But positing that creatures without conceptual capacities nonetheless use representations with indexical individual contents might seem implausible. Do we really need to posit that the perceptual contents of creatures like a mouse are that similar to first-person thought? An alternate view of perceptual representations suggests that we do not. Schellenberg (2016) and Mitchell (2021) argue that, instead, perceptual representations involve merely indexical locational contents.

Following Mitchell (2021), I will unpack the view that perceptual representations involve merely indexical locational contents using an idea from Evans (1982). Evans, in discussing the conceptual framework a subject uses to think about egocentric space, suggests that one locates objects in the environment relative to a perceptual map, whose coordinates are "given by the concepts 'up' and 'down', 'left' and 'right', and 'in front', and 'behind' " (p. 154). He further suggests that we understand the map's origin by analogy to the concept 'here': "we may call thinking about spatial positions in the framework centring on the subject's body 'thinking egocentrically about space'. A subject's 'here'-thoughts belong to this system: 'here' will denote a more or less extensive area which centres on the subject" (p. 154).

If how we conceptualize egocentric space centers (quite literally) on the use of the concept 'here,' then perhaps how we perceptually represent such space centers on a non-conceptual analogue of 'here.' Mitchell (2021) clearly articulates this idea when he proposes that "the non-conceptual analogue of 'here'... would not only be embedded within a holistic system of egocentric spatial notions, but would play a unique role within that system, namely referring to what Evans calls a *more or less extensive area which centres on the subject*" (p. 20). In other words, his thought is that in perception, one represents objects at locations relative to another location – that occupied by one's body – using indexical locational contents. This idea is anticipated by Schellenberg (2016). She argues that the origin of one's perceptual map can be represented using what she calls *de hinc* contents, which are indexical locational contents.

Both Schellenberg (2016) and Mitchell (2021) argue that this view of perceptual contents can be used to explain action. What is critical to such explanations seems to be "that one represents one's location in a dual mode: as both the point of origin of perception, and as the point of origin for bodily movement" (Schellenberg, 2016, p. 341). In other words, according to Schellenberg, the origin and axes of one's perceptual map must be the same as the origin and axes of one's actional map. Explanations of action require this match between one's perceptual and actional maps because it enables perceptual representations to inform action. But critically, that one's maps match in this way leaves open how one represents the maps' origins. As Schellenberg writes, "nothing in the requirement of having an origin and axes on the perceptual map requires any *de se* representation of the origin point via 'I'. All that is required is that there be some such origin and axes, and that they are available for perception and action to operate on as relevant locations and orientations" (p. 341). She thus proposes that the origin of one's perceptual map can be represented using *de hinc* contents. So long as the 'here' from which one perceives is the same as the 'here' from which one acts, one can use *de hinc* perceptual contents to offer explanations of action.

As we've seen, both views of perceptual contents seem to be able to explain a creature's actions. Given this, Mitchell (2021) argues that all else equal, we have reason to posit the use of merely indexical locational contents instead of indexical individual contents. Simplicity is supposed to push us to this conclusion. As he explains, "the *de se* reading implicates *complex* self-representational capacities in what are basic visual-spatial scenarios" (p. 26, emphasis added). Because the use of indexical locational contents can just as well explain a creature's actions, positing *de se* contents is "surplus to requirements" (ibid). Schellenberg (2016) likewise identifies the Locational Contents View as the simpler one. For example, she writes that "even quite sophisticated perceptual states have *mere de hinc* rather than *de se* content" (p. 341, emphasis added). While both authors seem to express similar ideas, exactly why the Locational Contents View is simpler remains somewhat opaque. I conjecture that the purported problem with the Individual Contents View is that it hyper-intellectualizes perception (see Burge, 2010). The concern is that the view treats perceptual representations as analogous to conceptual representations, even though perception is a more primitive cognitive system than thought. The Locational Contents View is thus thought to be preferable given that it avoids this problematic resemblance to thought and appears to equally well explain action.

1.2 The *Locational Contents View*'s motivation problem

I want to push back against the idea that our two views of perceptual contents have equal explanatory power. To show this, I first distinguish two desiderata of representational-level explanations of action. Such explanations should address how an action is spatially responsive to the environment and why a creature performs the spatially responsive action that it does. Following this, I argue that the view that perceptual representations have merely indexical locational contents satisfies the first desideratum but cannot easily satisfy the second. In the next section, I then argue that the view that perceptual representations have indexical individual contents also faces an explanatory challenge: while it easily satisfies the second desideratum, it struggles to satisfy the first. In this way, each of the two views captures something that the other leaves out.

To begin, I want to suggest that representational-level explanations of action should satisfy two desiderata insofar as they should be able to answer the following two questions:

1. *How is a creature's action spatially responsive to its environment?*

2. *Why does a creature perform the spatially responsive action that it does?*

Let's look at these questions one at a time. The first assumes that a creature's action will be "spatially responsive" to its environment and asks for an explanation of this. I use the term "spatially responsive" to describe actions that respond to the spatial features of a creature's environment. Such features are those that concern the locations, shapes, sizes, or movements of objects in the environment. A frisbee coming towards me is a spatial feature of my environment; the frisbee's color is not. If a frisbee is coming towards me and I reach my arm out to catch it, my action is spatially responsive. There is a match between the spatial features of my action and the spatial features of my environment.

Part of what explains the spatial match between one's action and one's environment is that the action is informed by perception. My action of reaching my arm out is spatially responsive because it is informed by my perception of the frisbee. In other words, it is because perceptual information about the frisbee's location is fed into my actional system that it can issue a motor command to reach for the frisbee. Plausibly, the way in which this perceptual information feeds into my actional system is indirect: it goes by way of higher-level cognition. But whether or not the route from perception to action is direct is orthogonal. What is important is simply that in order for one's action to be spatially responsive to one's environment, the former must be based, directly or indirectly, on perceptual information. In the absence of such information, one's actional system will lack information about the environment and so will not (except through luck) generate actions whose spatial features would match those in the environment.

I take it that most actions are spatially responsive to the environment. An action would not be spatially responsive if the link between a creature's perceptions and actions were severed. In this case, the action would appear random or disconnected from the context, unless sheer luck produced a spatial match between it and the environment. Infant movements provide a helpful example, as they often look random and so do not seem spatially responsive. The empiricist view holds that infants initially move "without regard to what they [see]" and gradually learn through trial and error to coordinate perception and action (Keil, 2013, p. 123). While now considered outdated, this view shows how a capacity for spatially responsive action could develop from increased coordination between one's perceptual and actional systems.[6]

The Locational Contents View can explain how an action is spatially responsive to one's environment. Recall that the view posits the use of a perceptual map, whose origin is represented using merely indexical locational contents. This entails that when, for example, I perceive the frisbee coming towards me, I visually represent it as approaching a location. While this location is in fact

[6]More recent research suggests that infant reaching movements are not as random as they seem, and that increased eye-hand coordination results from other developments in one's perceptual and motor systems, rather than development of the perceptual-motor link specifically (see Keil, 2013, pp. 123-125).

my location, it is represented using merely indexical locational contents and not indexical individual contents. Now, in order for my action of reaching for the frisbee to be spatially responsive to my environment, my action must be informed by perception. This will be possible so long as there is a coordinate transformation between my perceptual map and my motor command. A coordinate transformation allows for translations between coordinates used in different coordinate systems (see Gallistel, 1999). If my motor command to reach for the frisbee employs the same coordinate system as my perceptual representation of the frisbee, then any coordinate transformation from the latter to the former will be trivial. If the two maps differ, a coordinate transformation will enable my motor command to successfully target the frisbee's location. Either way, the Locational Contents View can explain how my perception informs my action. As Schellenberg (2016) expresses, all that matters is that there be "some such origin and axes, and that they are available for perception and action to operate on as relevant locations and orientations" (p. 341). That the origin of my perceptual map is represented using merely indexical locational contents does not seem to impair the link between perception and action. As of yet, we find no need to posit indexical individual contents.

Let's now turn to consider the second desideratum of representational-level explanations of action. This asks why a creature performs the spatially responsive action that it does. Why, for example, do I reach my arm out to catch the frisbee rather than moving my head towards it? Both actions could be spatially responsive to my environment insofar as they could be informed by the same perceptual inputs, and yet I perform one action and not the other. An account of action should explain why.

One way to explain why a creature performs one action and not another is to appeal to its motivations. To unpack this idea, I appeal to belief-desire explanations, which hold that actions are caused by a combination of beliefs and desires.[7] In such explanations, it is the desire component

[7]For discussions of belief-desire explanations of action, see Davidson (1963) and Fodor (1987). I adopt belief-desire explanations for their simplicity. While these accounts can be expanded to include intentions (see Bratman, 1987; Velleman, 1992), these elaborations do not affect my central claims concerning the role of indexical perceptual contents in action explanation. What is important is only that the explanation considered is reasons-based rather than

that provides the motivational force behind the creature's action. Consider our frisbee example: a simple belief-desire explanation of my action of catching the frisbee might take it to be caused by my perceptual belief about the frisbee's location and my desire that I play frisbee with my friend. It is because my desire motivates me to catch the frisbee, rather than collide with it, that I perform the former action and not the latter. While providing this kind of explanation might seem trivial, I will argue that it poses a significant challenge for the Locational Contents View.

Trying to explain my action of catching the frisbee in belief-desire terms helps illustrate why the Locational Contents View faces difficulties answering our second explanatory question. Let's first consider what the relevant belief might be. Plausibly, it is a perceptual belief with roughly the same contents as the perceptual representation on which it is based. If so, then my belief, like my perceptual representation, carries indexical locational contents and so refers to a particular location, not an individual. Let's now turn to my desire. We can imagine that having gone to the park with some friends, I desire that I play frisbee with them – that I toss the disc to my friends when it's in my possession and that I catch it when it's thrown to me. As can be seen, this desire is *de se*; it indexically refers to my self. The problem is that there is a disconnect between my perceptual belief and my desire. Because the former merely refers to a location while the latter refers to my self, the two cannot connect to rationally explain my action. We can further understand this by recasting Perry's example of the messy shopper in belief-desire terms. Recall that Perry does not rearrange his sugar sack when he has the *de dicto* belief '*The shopper with the torn sack is making a mess*.' Nonetheless, we can imagine that Perry desires that he be considerate to others. So why don't his belief and desire prompt him to start rearranging his sugar sack? Because the two do not connect: his belief is *de dicto* while his desire is *de se*. The idea is that a similar disconnect

---

merely causal.

One might wonder whether my arguments extend to non-human animals, given skepticism about whether belief-desire models apply to them. What matters for my purposes is simply that an animal's motivation to act connects to its perceptual representations. This connection is preserved across a range of alternative frameworks used to explain animal behavior, including those involving proto-beliefs and proto-desires (Bermúdez, 1998, pp. 118–123), non-conceptual intentions (Kaufmann, 2015), and perceptual representations with normative content (Sebo, 2017). So, I expect that my arguments could be re-framed so as to apply to non-human animals.

appears in our frisbee example between my perceptual belief and my desire.

To be clear, the Locational Contents View only concerns the contents of perceptual representations. The problem raised here does not concern these contents per se, but rather suggests that they are insufficient for explanations of action. More specifically, the problem is that when perceptual representations with mere indexical locational contents are fed into belief, resulting belief-desire explanations are unsuccessful.

Changing the kind of action explanation offered will not address this problem. Mitchell (2021), for example, attempts to explain a similar action – moving to avoid a ball – in terms of the contents of visual experience and intentional action. Using 'x' to stand for the ball and '*h*' to stand for what I have called indexical locational contents, he claims that "the content of the visual experience is ⟨x moving towards *h*⟩, and that the content of the intentional action could be ⟨move left of *h* to avoid x⟩" (pp. 24-25). Mitchell further explains that "taking the content of [one's] visual experience at face value provides [one] with a non-inferential reason for *moving*" (p. 24). But this is where we disagree. The content of one's visual experience does *not* provide one with a reason to move because the content does not refer to one's self. The content merely specifies that an object is approaching a location, and that, in itself, does not give one reason to do anything. To frame the issue differently: while it is plausible that one has a standing desire not to be hit by flying objects, it is implausible that one has a standing desire that flying objects not reach a particular location. In other words, the motivation that could plausibly explain why one moves does not connect with the content of one's visual experience as Mitchell characterizes it. Without reference to the self, perceptual contents cannot link up with the relevant motivations to explain action.

In the face of this problem, one might propose that the functional connection between perception and action guarantees that specific perceptual inputs cause specific intentional actions. Using Mitchell (2021)'s example, the idea would be that one's perception-action link is such that when one has a visual experience with the contents ⟨x moving towards *h*⟩, one automatically forms an intention with the content ⟨move left of *h* to avoid x⟩. If this input-output pair were hardwired into the connection between one's perceptual and actional systems, it would causally explain why one moves out of the ball's path. The broader idea behind this action explanation is that one performs the spatially responsive action that one does because of how one's perceptual representations, which involve merely indexical locational contents, are processed by one's actional system.[8]

The problem with invoking functional architecture in this way is that it does not allow actions to be context sensitive. To see why, we should first note that it is not plausible that all possible pairs of perceptual inputs and actional outputs are built into the functional connection between perception and action. There are simply too many pairs. Given this computational constraint, any function from perceptual inputs to actional outputs will need to generalize away from the specifics of the perceptual scene. With this in mind, we might imagine that one's perception-action link operates such that every time one sees a ball coming towards one, one moves out of its way. But, of course, it is not the case that one always acts in this way. Actions are context-sensitive. When a ball is coming towards me, there are many actions I might take. I might move out of its path, or jump to catch it, or wait until it reaches me to catch it, and so on. Positing a fixed relation between my perceptual inputs and actional outputs would preclude this flexibility.

A different approach to defending the use of the Locational Contents View in action explanations might introduce an additional belief to explain why one is motivated to act as one does. Recall that in the frisbee case, the challenge is that my perceptual belief, '*A frisbee is approaching here*,' does not seem to connect to my *de se* desire that I play frisbee with my friends. The new suggestion is that a complete belief-desire explanation of my action requires more than just a perceptual belief. Such an explanation must also include the belief '*I am here*,' which combines with my perceptual belief to yield the belief '*A frisbee is approaching me*.' In contrast to my perceptual belief, this *de se* belief could connect to my *de se* desire that I play frisbee. The combination of

[8]It is tempting to think that this action explanation simply mis-describes the perceptual contents that it posits. While insisting that perceptual representations have merely indexical locational contents, this account simultaneously proposes that perceptual representations function to motivate actions. One might argue that this is a mistake. Perceptual representations that motivate actions are properly described as involving indexical individual contents. One cannot erase such contents by packing their functional import into the perception-action link.

my *de se* belief and my *de se* desire would explain why I catch the frisbee.

While the addition of a bridge premise would resolve the Locational Contents View's motivation problem, it would also tarnish the view's purported simplicity. As discussed, proponents like Schellenberg (2016) and Mitchell (2021) favor this view because it avoids positing seemingly sophisticated contents – indexical individual contents – in basic perceptual scenarios (see §1.1). But this simplicity is preserved only by shifting the complexity elsewhere – namely, by requiring an additional *de se* belief and inference in our explanations of action. In effect, the Locational Contents View avoids hyper-intellectualizing perception only to hyper-intellectualize the perception-action link. This trade-off arguably leaves our action explanations worse off: positing an additional *de se* belief and inference in our action explanations may be more complex than simply accepting that perception involves indexical individual contents.

To wrap up, I have argued that the view that perceptual representations have indexical locational contents has difficulty satisfying one of the two proposed desiderata of representational-level explanations of action. Specifically, it struggles to offer a simple explanation of why one performs the spatially responsive action that one does. This constitutes a serious shortcoming.

1.3 The *Individual Contents View*'s coordination problem

I now turn to evaluate whether the Individual Contents View satisfies our two desiderata of representational explanations of action. I start by showing how the view easily explains why one performs one action and not another, and so satisfies the second desideratum. I then explore the challenges that arise when the view attempts to satisfy the first desideratum and explain how one's action is spatially responsive to one's environment. These challenges give us reason to look for a new view, which I offer in §1.4.

The second question that I claimed representational-level explanations of action should answer is why one performs the spatially responsive action that one does. Why, for example, do I reach out to catch the frisbee rather than bonk my head into it? The Individual Contents View can be used in belief-desire explanations to easily answer this question. In broad strokes, the idea is that one performs the action that one does because one's *de se* perceptual belief connects with a *de se* desire that motivates the action one performs. With respect to our frisbee example, the explanation goes as follows. First, I visually represent the frisbee using indexical individual contents. Next, I form a perceptual belief whose content roughly matches that of my visual representation. My belief is thus '*A frisbee is approaching me*.' This perceptual belief connects with a *de se* desire that I play frisbee with my friends. Lastly, these two states – my belief state and my motivational state – cause my action. The reason that I catch the frisbee rather than bonking my head into it is that my desire motivates the former action and not the latter.[9]

The Individual Contents View can be used in belief-desire explanations that are considerably simpler than ones that use the Locational Contents View. As discussed, belief-desire explanations that use the Locational Contents View require that one infer a *de se* belief from a perceptual belief (that carries merely indexical locational contents) and the bridge premise '*I am here*.' It is the inferred *de se* belief that connects to the desire that motivates action. But, in belief-desire explanations that use the Individual Contents View, no inference or bridge premise is needed. This is because one's perceptual belief is already *de se* and so can itself connect with the desire that motivates action.

I've argued that the Individual Contents View provides a simple account of why one performs the spatially responsive action that one does. In this way, it satisfies the second desideratum of representational-level explanations of action. But how does the view fare with respect to the first desideratum? Can it explain how a creature performs an action that is spatially responsive to

[9] Answers to more detailed questions about why one performs the spatially responsive action that they do will likely refer to motor planning that happens downstream of the formation of belief and motivational states. For example, we might wonder why I catch the frisbee with my right hand rather than my left. The suggestion is that this is caused by motor planning that occurs after my intention to catch the frisbee. My actional system determines the best – perhaps the most energy efficient – way to catch the frisbee. In this case, that might be to use my right hand. While these details are part of a more complete explanation of why I perform the action that I do, they are not directly relevant to or affected by the belief-desire explanation offered by the Individual Contents View. The role of the Individual Contents View in representational-level explanations of action concerns the relations between one's perceptual, belief, and motivational states and how these states prompt action.

its environment? I first show that to answer this question, we must assume that the indexical individual contents described by the view refer to a bodily self. I then argue that having interpreted indexical individual contents in this way, the view nonetheless encounters numerous challenges in explicating how these contents can inform spatially responsive actions.

To illustrate why indexical individual contents must refer to a bodily self in order to explain a creature's capacity for spatially responsive action, let us suppose the opposite for the purposes of argument. Suppose that these indexical individual contents refer to a Cartesian ego. If so, then my visual representation of the frisbee carries indexical individual contents that refer to my Cartesian ego and not my body. Of course, this raises an immediate difficulty: my perceptual representation of the frisbee is not only hard to interpret but also at risk of involving a category mistake. If my ego lacks spatial properties, it cannot be properly described as spatially related to a frisbee. My ill-formed visual representation thus fails to locate the frisbee in my environment. As such, the representation cannot be used to explain why my (body's) action is spatially responsive to its environment. If I were to miraculously catch the frisbee, this feat could not be explained by referencing my visual representation. The view that the indexical individual contents used in perception refer to something like a Cartesian ego is clearly a non-starter.

We can articulate this idea slightly differently by further investigating the notion of a perceptual map. As discussed, the objects of perception are thought to be represented on a perceptual map used to specify spatial relations (§1.1). One relatum of these spatial relations is the objects of perception. These objects have spatial features: they have sizes, shapes, and occupy locations. At issue in the debate about perceptual contents is what the other relatum is. One possibility, offered by the Locational Contents View, is that it is a spatial location. Another possibility, posited by the Individual Contents View, is that it is the perceiver. Importantly, this latter possibility seems to presuppose that the perceiver – one's self – has spatial features: it is the kind of thing that can stand in spatial relations to other objects. In this way, the Individual Contents View presupposes that the self is similar to the objects of perception. Just as those objects have sizes, shapes, and occupy locations, the self has a size, shape, and occupies a location.

It thus seems that the Individual Contents View takes indexical individual contents to refer to a *bodily* self. More specifically, such contents refer to the embodied creature who tokens a perceptual representation with such contents. Cassam (1997), who takes spatial perception to be self-locating, articulates this idea when he writes that "in experiencing objects as spatially related to one, one literally experiences the *bodily* self as located in the perceived world" (p. 53, emphasis added). He reiterates this idea as follows: "for it is one's Body that is at the point of origin of egocentric space, and in relation to which other bodies are experienced as being to the left or right, above or below" (p. 53).

Having interpreted indexical individual contents to refer to a *bodily* self, let me clarify what perceptual contents look like according to the Individual Contents View I am articulating. This view holds that perceptual representations specify spatial relations between perceived objects and the perceiver using indexical individual contents that refer to a bodily self. Crucially, such representations carry only indexical individual contents and not also indexical locational contents. In this way, perceptual representations specify relative spatial relations and presuppose that the relata have absolute locations, though these locations are not included in the contents themselves.

It remains unclear whether this view aligns with those who describe perceptual representations as self-locating. Consider Alsmith (2017):

> "An experience represents its subject's location when its content includes a spatial relation between the subject and the object of the experience. It includes a relation of the kind *aRx*, where *a* refers to the object of experience, *x* to the subject of the experience, and *R* to the spatial relation between them. In representing *R*, the content specifies the relative location of each relatum as represented by the experience" (p. 267).

That perceptual contents include "a relation of the kind *aRx*, where *x* is the subject of experience" suggests these contents do *not* include indexical locational contents but rather include merely indexical individual contents. Yet, Alsmith simultaneously claims the perceptual experience represents the subject's *location*. I highlight this tension only to illustrate the interpretive challenges in understanding the common view that perceptual contents are self-locating. Regardless of what proponents of this view maintain, the Individual Contents View I am considering here takes a clear position: perceptual contents include only indexical individual contents.

Now, let us consider how indexical individual contents might be used to explain spatially responsive action. Recall that explanations of spatially responsive action seem to require that perception informs action. Further recall that perception can inform action only if there is a coordinate transformation between one's perceptual and actional maps. The Individual Contents View assumes that the origin of one's perceptual map is represented using indexical individual contents that refer to one's bodily self. So long as this origin can be coordinated with the origin of one's actional map, then one's perceptual representations can inform one's actions. It seems reasonable to assume that the origin of one's actional map will also be represented using indexical individual contents that refer to one's bodily self. Thus, coordination between perception and action should be straightforward.

But the fact that perception and action use *many* spatial maps poses a problem for this explanation. Evidence suggests that human visual perception alone employs retinotopic, cyclopean, and head-centered reference frames.[10] Not only does each of these reference frames play a critical role in visual perception, but also, each of their origins corresponds to a different location. The Indexical Contents View would suggest that each origin is represented using indexical individual contents. But if this is right, then the origins of these maps won't be distinguished from one another and so, coordination between these maps and those used in action will be unsuccessful. To illustrate the problem, imagine that I perceive the frisbee and visually represent it relative to

[10]For evidence of retinotopic representations, see Wandell & Winawer (2011); for cyclopean representations, see Julesz (1971) and Ono et al. (2002); and for head-centered representations, see Galletti et al. (1993) and Sun & Goldberg (2016). Other perceptual modalities also appear to use multiple reference frames. For instance, evidence suggests that auditory perception employs both eye-centered and head-centered reference frames (Groh, 2014; Mullette-Gillman et al., 2005).

my cyclopean reference frame, whose origin corresponds to the midpoint between my eyes. The frisbee might be represented as 20 degrees to the right, 15 degrees above, and 12 feet away from this point. But, if the origin of my cyclopean reference frame is not distinguished from the origin of my head-centered frame, my actional system won't be able to determine whether these coordinates are relative to the midpoint between my eyes or the center of my head. In short, my visual representation won't be of use to my actional system.

One might try to sidestep this problem by distinguishing between subpersonal and personal-level perceptual representations. The thought would be that while one's subpersonal perceptual representations make use of many different spatial maps, one's personal-level perceptual representations do not. Perceptual experiences appear unified, strongly suggesting the use of only one map at the personal-level. If only one map is used, then there is no need to distinguish its origin to make it fit for coordination with action.

It is not obvious that personal-level perceptual representations are unified on a single perceptual map. Perhaps, the unity of perceptual experience is instead achieved through the coordination of many maps.[11] But even if we grant that personal-level perceptual representations are unified on a single perceptual map, this response has its drawbacks. First, it restricts the scope of the Individual Contents View to only apply to personal-level perceptual representations. Second and more importantly, it raises a new challenge. This is to specify more precisely which part of the bodily self is referred to by the origin of one's single and unified personal-level perceptual map. That it might be the whole body seems implausible. The body occupies a significant volume and its parts are distributed across different locations. This entails that specifying spatial relations relative to the whole body will be imprecise. Representing a frisbee as 20 degrees to the right,

[11] Skepticism towards the idea that personal-level perceptual representations are unified on a single perceptual map seems present in Briscoe (2009). Seemingly referring to the use of many egocentric reference frames in subpersonal processing, he writes that "this subpersonal representational arrangement seems to be reflected at the person level. When I see an object's egocentric location, I do not simply see its location relative to myself. Indeed, there is no privileged point in (or on) my body that counts as *me* for purposes of characterizing my perceived spatial relation to the object" (p. 425). Alsmith (2021) explicitly argues for the related claim that bodily experience and action can be explained without positing "a representation of the body as an integrated whole" (p. 2195).

15 degrees above, and 12 feet away from the whole body is under-specified: it is unclear where these coordinates are measured relative to. This suggests that for perceptual representations of spatial relations to be precise, they must be specified relative to an origin that corresponds to a single point. But which bodily point is referred to by the single, unified perceptual map posited? The obvious candidates are the center of one's head and the center of one's torso. Alsmith (2017) articulates the advantages of these two origin points as follows: "The head houses a great number of sensory organs that are particularly significant for spatial representation, namely the eyes, ears, and the vestibular labyrinth. Yet, ... the torso is the stable continent relative to which other parts are mere peninsulas and hence the most likely point of reference for the construction of a consistent egocentric representation" (p. 273). Given these competing factors, it's not clear why a proponent of the Indexical Contents View would choose one point on the body over the other.

To wrap up, I have argued that the Individual Contents View can explain why one performs the spatially responsive action that one does, and, thus, the view satisfies the second desideratum of representational-level explanations of action. Where the view encounters difficulty is in addressing the first desideratum: it struggles to explain how one performs spatially responsive actions in the first place. More specifically, problems arise when trying to fit the idea that indexical individual contents refer to a bodily self with the idea that such contents serve as the origin of one's perceptual map(s).

1.4 The *Located Individual View*

Our two views of perceptual contents have different explanatory strengths and encounter different explanatory problems. The Locational Contents View can be used to explain *how* one performs a spatially responsive action but cannot be used to easily explain *why* one does so. In short, it struggles to answer questions about a creature's motivations. By contrast, the Individual Contents View can be used to explain *why* one performs a spatially responsive action but cannot be used to easily explain *how* one does so. It thus struggles to answer questions concerning the coordination between perception and action. In this section, I propose to integrate the two views and so combine their explanatory strengths.

According to what I call the Located Individual View, perceptual representations carry both indexical individual contents and indexical locational contents. These refer to a perceiver and a spatial location (that is occupied by the perceiver), respectively. We might call the conjunction a "located individual." Strictly speaking, this is a view about the contents of perception. But to get a fuller picture of how such contents might be used within our perceptual systems, consider how they could be represented. Suppose, for example, that the origin of a perceptual map – such as the cyclopean reference frame – is explicitly represented in vision. I would like to suggest that this origin representation, o, has the content $\langle \text{i, h} \rangle$. That is, the origin carries a pair of indexical individual and locational contents. When I token o, it simultaneously refers to my self and a specific location – the midpoint between my eyes. It thus represents what I have called a located individual.

These two indexical contents serve complementary roles. Indexical locational contents establish a spatial framework while indexical individual contents identify who is located at its origin. The thought is that neither content alone captures the nature of egocentric spatial perception. The use of merely indexical locational contents privileges a spatial point without anchoring it to a perceiver. The use of merely indexical individual contents refers to a perceiver without situating her in space. Together, these contents capture that perception presents the world as spatially structured relative to one's self.

The Located Individual View seems to resemble some articulations of the idea that perceptual contents are "self-locating."[12] For example, Schwenkler (2014) presents what he calls the

[12] One articulation of the self-locating view that clearly resembles the Located Individual View appears in Mitchell (2021). While Mitchell himself argues against the self-locating view, he explicitly construes it as positing that perceptual contents "index the subject relative to a location" (p. 25). Peacocke (1999) also seems to express a view like the Located Individual View. He claims that perceptual contents can have the form 'I am *F*' and "[represent] the subject as having some location in the spatio-temporal world" (p. 265). Thus, perceptual contents appear to include both indexical individual and locational elements. Perhaps, the same is true of creatures at Degree 1. Such creatures are characterized by a capacity for non-conceptual *de se* representation, but Peacocke (2014) also describes them as having the resources to think through "their own past path through space" (p. 36).

“Self-Location Thesis,” which claims that “simply in virtue of its perspectival character, visual experience can include the *location* of the perceiver among its face value contents” (p. 3, emphasis added). At first glance, one might interpret this as suggesting that perceptual representations carry mere indexical locational contents. However, other parts of Schwenkler’s discussion clarify that this is not his intended view. He writes, for example, that one can represent something as “to *my* right” or “to *my* left,” indicating that perceptual representations also include indexical individual contents (p. 7). While his view and the Located Individual View thus appear similar, his discussion somewhat obscures the dual role of indexical locational and individual contents. Explicitly distinguishing between these two types of indexical contents not only can clarify our understanding of perceptual contents, but also can illuminate how they contribute to action explanations.

The Located Individual View can address both explanatory questions about action. Regarding the first, the view can be used to explain how one’s actions are spatially responsive to one’s environment. As we’ve seen, spatially responsive actions require perception to inform action. This is possible given the Located Individual View, which takes the origin of a perceptual map to carry indexical locational contents. So long as locations given relative to this origin can be translated into locations on one’s actional map, there will be a link between perception and action that enables the performance of spatially responsive actions. This explanation is the same as that provided by the Locational Contents View.

The Located Individual View supports action explanations with broader explanatory scope than the Individual Contents View. This is because explanations using the former view succeed regardless of the number of perceptual maps used. According to the Located Individual View, each perceptual map is distinguished by its origin, which carries the content $\langle$i, h$\rangle$. While indexical individual contents refer to the same perceiver across maps, indexical locational contents refer to different bodily locations. Thus, the origin of each map refers to a different located individual. Because one’s perceptual maps are distinguished from each other in this way, coordinate transformations can be used to translate between them and one’s actional maps. This entails that the Located Individual View can support both subpersonal and personal-level explanations of action.[13] This contrasts with the Individual Contents View. As discussed, the Individual Contents View cannot distinguish between multiple maps within a single perceiver, which limits its use to personal-level explanations (§1.3).

Turning to the second explanatory question, the Located Individual View can be used to explain why one performs the action that they do. We can see this by observing how the view is compatible with belief-desire explanations. According to the view, one's perceptual representations carry indexical individual contents. Thus, these representations can inform *de se* perceptual beliefs. These beliefs, in turn, can connect with one's *de se* desires that motivate one to act as one does. We have already seen this explanation: it is the same as that offered by the Individual Contents View (see §1.3).

Now, there is no need to posit use of the bridge premise '*I am here*' to connect one's perceptual belief with one's motivational state. One's perceptual belief already carries these contents because it is based on a perceptual representation that carries indexical individual and indexical locational contents. Nor is there need to posit a specific kind of functional architecture that might fix the relation between perceptual inputs and actional outputs. An explanation of why one acts as one does is fully available at the content level. In these ways, the Located Individual View's answer to the second explanatory question avoids the difficulties encountered by the Locational Contents View.

1.5 Conclusion

In conclusion, I have argued that explanations of action seem to require that perceptual representations carry both indexical individual and indexical locational contents. The former contents help explain how one is motivated to perform the action that one does; the latter contents help explain

[13] By accommodating the use of multiple perceptual maps, the Located Individual View also avoids the need to identify a single bodily point as primary in perceptual representation. This marks another distinction between it and the Individual Contents View, which must determine whether the origin of one's single perceptual map corresponds to the center of the head or torso.

how one's perceptual representations are coordinated with one's actional representations. Views which posit that perceptual contents include only one of these indexical elements encounter significant challenges when used in action explanations. To overcome these challenges, I have proposed that perceptual representations carry both kinds of indexical contents. More specifically, I have suggested that each origin of a perceptual map represents a "located individual," that is, an individual and a specific location occupied by that individual. Together, these contents capture that perception presents the world as spatially structured relative to one's self.

Chapter 2

Egocentric Reference Frames in Perception

Perceptual experiences are egocentric. Put roughly, this means that one experiences the objects of perception from one's own point of view. As I type, I'm not simply looking at my computer. I'm looking at my computer from a particular point of view, which affects how and where the computer appears relative to me. It currently appears head on and three feet in front of me. If I move forward and to the left, its side will appear to my right. In this way, my perceptual experiences are centered on me: they place me "as a subject at the center of [my] world" (Campbell, 1994, p. 6).

Philosophical and empirical accounts of perception often capture its egocentric nature by positing that perception involves the use of egocentric reference frames. These are frameworks used to specify the locations of perceived objects relative to the perceiver. Consider the cyclopean reference frame used in early visual perception. This frame specifies spatial relations relative to the perceiver's eyes using a coordinate system whose origin corresponds to the midpoint between the eyes and whose axes correspond to the front/back, left/right, and up/down directions relative to the eyes. When I visually represent my computer, I specify its orientation and location in my cyclopean reference frame. As I move my eyes, the frame shifts – the midpoint between my eyes and directions of my eyes move – and so I come to represent the computer at a new orientation and location. It is because the frame is context-sensitive in this way that it seems to capture my visual point of view.

More generally, egocentric reference frames are what make perceptual representations representations from an individual's point of view. They are, if you will, the representational glue between objects in the environment and the perceiving self. Despite this critical role, questions about their nature and use remain underexplored. One central issue concerns what these frames contribute to the content of perceptual representations. Although I have suggested that they might carry both individual and locational contents (§1.4), a systematic account of this proposal remains to be developed. A related but distinct question concerns what it means to say that such frames are "used" in perception. Are egocentric reference frames constituents of complex perceptual representations or are they built into the architecture of the mind?

These questions take on particular urgency when we consider how perceptual representations must function in action explanations. In the previous chapter, I argued that successful action explanations require that perceptual representations carry both indexical individual contents and indexical locational contents – what I called the Located Individual View. If egocentric reference frames are indeed the mechanism underlying the use of these two indexical contents, then understanding their nature and operation becomes central to a complete theory of how perception guides action.

This chapter advances a novel set of claims to address these questions. After providing additional background on egocentric reference frames (§2.1), I elaborate a view according to which these frames contribute two kinds of indexical contents to perception: indexical individual contents (which index to the perceiver) and indexical locational contents (which index to a location) (§2.2) – thereby supplying a mechanism by which the Located Individual View might be realized. I then argue that when used in their standard role as anchors from which to specify spatial relations, egocentric reference frames remain implicit: they contribute to the accuracy conditions of perceptual representations without themselves being explicitly represented (§2.3). Lastly, I propose that these frames are best understood as encoded in the functional architecture of perceptual systems (§2.4). It is in virtue of this that egocentric reference frames shape the accuracy conditions of perceptual representations without being explicitly represented. In building on elements of these claims anticipated by different philosophers in the literature, this chapter develops a unified account of how egocentric reference frames operate in perception. Taken together, this account entails that indexical individual information is implicit and encoded in the functional architecture of our perceptual systems. I briefly discuss this consequence in §2.5.

2.1 Introducing egocentric reference frames

We can begin to understand what egocentric reference frames are by looking at the genus *reference frame* of which they are a species. A reference frame is a framework for specifying spatial locations. Groh (2014) helpfully unpacks this idea by considering how we talk about space (p. 161). Imagine, for example, that I describe the spatial arrangement of items on my desk by specifying their locations relative to my computer. Here, I am using a "computer-centered" reference frame because my computer functions as the origin of a framework for specifying spatial locations. Alternatively, I might (painstakingly) attempt to specify these locations relative to the nearest library, thereby using a "library-centered" reference frame. While more or less practical, any one of an infinite number of points could function as the origin of a reference frame used to specify the locations of items on my desk.

In paradigmatic cases, a reference frame specifies locations using a *coordinate system*. A *polar coordinate system* specifies a location l in 3D-space by i) the distance between l and a reference point or *origin*, and ii) the angles between l and two reference directions or *axes*.[1] One angle is given relative to an up/down (i.e. zenith) direction and the other angle is given relative to a left/right (i.e. azimuth) direction. While polar coordinates can be translated into Cartesian coordinates and vice versa (Klatzky, 1998, p. 5), I use polar coordinate systems throughout this chapter: they mirror how we naturally describe the locations of objects using distance and direction pairs.[2]

[1]I set aside analysis of reference frames that use coordinate systems that likely lack an origin and/or axes. Tactile perception, for example, involves the use of a somatosensory map (see Medina & Coslett (2010)'s discussion of primary somatosensory representations). This kind of map does not seem to require an origin: we can imagine that each location on the skin is specified by a unique address, none of which is privileged.

[2]Gallistel (1990) notes that although one can translate between polar and Cartesian coordinate systems, there are

Campbell (1994) writes that the use of an *egocentric reference frame* involves "thinking about the space [one is in] from a particular point of view, as a subject at the center of one's world" (p. 6). Imagine Bo – a retriever – is lying down and staring at a ball in front of him. He is perceiving the ball "from a particular point of view," namely, from his *own* point of view. We can begin to offer a functional account of how he perceives from his own point of view by positing that he uses an egocentric reference frame. In other words, Bo specifies the locations of objects in the environment using a reference frame whose origin corresponds to a point on his body.

The use of an egocentric reference frame is contrasted with that of an *allocentric reference frame*, in which one thinks about the space "independently of any particular viewpoint on it, in an impersonal or absolute way" (Campbell, 1994, p. 6). Allocentric frames are characterized 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). Using a library-centered frame is an example of the former; using a map of the New York City transit system is an example of the latter. While they differ with respect to the role of an origin, both seem "impersonal" insofar as they map spatial relations independent of any particular perceiver.

As discussed, human visual and auditory representations are used with an egocentric reference frame. But this is not to say that a given perceiver uses only one such frame. Different perceptual representations are used with different reference frames. Empirical research on the use of reference frames in perception attempts to determine which reference frame is used with a given perceptual representation by observing which changes to the spatial relations between a stimuli and a perceiver affect the perceiver's perceptual representation of the stimuli. A toy example helps to illustrate this experimental paradigm. Imagine that while looking at the computer in front of me, I hear the constant sound of rain falling outside the window to my left. If I shift my gaze, my auditory experience will not change. The same is true if I move one of my hands or feet. The only bodily

---

computational reasons to posit use of the latter (p. 76). The choice between coordinate systems, however, does not affect the discussions to follow, which focus on the role of a coordinate system's origin and axes.

movements that will affect my auditory experience are ones that involve changes to the position of my head. If, for example, I turn my head and look directly at the window, the rain will sound louder. This suggests that my auditory representations specify locations relative to my head. From this, we can posit that my auditory representations are used with a head-centered reference frame.

For simplicity, my discussion to follow largely focuses on only one egocentric reference frame: the cyclopean reference frame. Positing the use of this frame is thought to solve a puzzle regarding visual perception. The puzzle is that we receive two sources of visual information in the form of two 2-D retinal images, one from each eye, but we experience one 3-D visual percept. This suggests that our two retinal images are fused together. Some researchers posit that information from both eyes is integrated relative to a single reference frame – the cyclopean reference frame – that takes the midpoint of the eyes as its origin (Barendregt et al., 2015; Julesz, 1971; Ono & Barbeito, 1982; Ono et al., 2002; cf. Erkelens & van Ee, 2002).

2.2 The dual contents of egocentric reference frames

While the foregoing account explains how egocentric reference frames enable perception from one's own point of view, it leaves open what their contents are. In this section, I address this lacuna by developing what I call the *Dual Contents View*, according to which egocentric reference frames carry both indexical individual contents and indexical locational contents. I begin by offering an account of the accuracy conditions of perceptual representations that takes them to involve both the perceiver and her location. After discussing how this account aligns with the Located Individual View developed in Chapter One, I suggest that the perceiver and her location appear in these accuracy conditions in virtue of the use of egocentric reference frames. This entails that egocentric reference frames carry contents corresponding to both the individual and the location. I then develop this *Dual Contents View* using a Kaplanian framework and motivate it by showing how it explains the two functions of egocentric reference frames.

Consider a case where one visually represents some object *o* in their environment. As discussed, there is strong empirical support for the idea that in early visual processing, one represents perceived objects in their cyclopean reference frame, whose origin corresponds to the midpoint between the eyes and whose axes correspond to directions relative to the eyes. Thus, further suppose that object o is located at distance d from the midpoint between the perceiver's eyes and an angle θ from directions relative to their eyes. Given this setup, I propose that their perceptual representation R in context c will have the following accuracy conditions:

(1) R is accurate at world w and context c iff object o is located at distance d and angle θ from location l(c) and l(c) is the midpoint between the eyes of individual i(c) in w.

These accuracy conditions contain two parts connected by a conjunction. The first part takes R to be accurate if o is a certain distance and direction from a location in the context. Thus, R's accuracy depends on a set of spatial relations holding between the object perceived and a location. This should be intuitive. After all, the perceptual representations that we are concerned with are spatial representations that are anchored to a particular location. Note, also, that this anchoring location l is context-sensitive. In this way, the accuracy conditions of a representation tokened at one location will necessarily differ from the accuracy conditions of a representation tokened at another location.

The second part of R's accuracy conditions mark a second way in which R's accuracy conditions are context-sensitive: they are sensitive to the individual in the context. Thus, the accuracy conditions of a representation tokened by one individual will necessarily differ from the accuracy conditions of a representation tokened by another individual. This captures the idea that a point of view is partially distinguished by the individual whose point of view it is. A similar idea is sometimes articulated in terms of the "for-me-ness" of perceptual experiences: they are for me, or mine, in a way that distinguishes them from the perceptual experiences of others (see Zahavi & Kriegel, 2015). If this is right, there is reason to think that at the representational level, the individual must figure in the accuracy conditions of her perceptual representations, thereby distinguishing the perceptual representations of one individual from those of another.

The second part of R's accuracy conditions posits a specific relation between the anchoring location and the individual in the context, namely, that the midpoint between the eyes of the individual occupies the anchoring location. This condition assumes that some relation holds between bodies and individuals. What exactly this relation is is not relevant for my purposes, but I will outline two possibilities. One is a Cartesian view according to which the individual stands in some kind of ownership relation to her body. In this case, discussion of an "individual's eyes" indicates that a Cartesian ego owns a particular set of eyes. Another option is to think that individuals are fundamentally embodied. In this case, an "individual's eyes" refers to one part of a physically embodied individual. All that matters for present purposes is that there is a one-to-one correspondence between bodies and individuals so as to make interpretable phrases such as "the eyes of an individual."

Specifying which part of the body occupies the anchoring location allows perceptual representations used with different reference frames to be distinguished from each other. To see how, imagine that *R*'s accuracy conditions merely specified that part of the individual's body occupied the anchoring location. In this case, *R* would be accurate were the perceived object at the distance and angle represented from *any* point on the individual's body that occupied the anchoring location. For example, *R* could be accurate regardless of whether the midpoint between the perceiver's eyes or the center of her body occupied the anchoring location. As a result, it would be impossible to distinguish *R* when used with the cyclopean reference frame from a nearly identical representation used with any other egocentric reference frame. Specifying which bodily point occupies the anchoring location thus is crucial for distinguishing perceptual representations used with different egocentric reference frames.

That both a location and an individual figure in the accuracy conditions of perceptual representations comports with what I called the Located Individual View, discussed in the previous chapter. As I argued there, perceptual representations must carry both indexical individual contents and indexical locational contents to explain action. The former help explain why perceptual states can motivate action by connecting to the perceiver's desires and goals; the latter help explain how perceptual states can inform action by structuring a reference frame. If a subject and a location appear in the contents of perceptual representations, then presumably, both should also figure in the accuracy conditions of perceptual representations.

The accuracy conditions I have described point towards a semantic interpretation of egocentric reference frames that I call the *Dual Contents View*. According to this view, egocentric reference frames carry two kinds of indexical contents: indexical locational contents and indexical individual contents. This view emerges from considering which elements in the accuracy conditions of a perceptual representation should be attributed to the egocentric reference frame with which it is used. The object o, distance d, and direction θ presumably appear in R's accuracy conditions in virtue of constituent representations of an object, distance, and direction.[3] This leaves the location l and midpoint between the eyes of individual i unaccounted for. The *Dual Contents View* suggests that these elements figure in R's accuracy conditions in virtue of the use of the cyclopean reference frame.

I will use a modified version of Kaplan (1989)'s theory of indexicals to further explicate the *Dual Contents View*. For background, Kaplan takes indexicals to have two kinds of meaning: a *character* and a *content*. The character of an indexical is a rule that determines the content in any given context. So, for example, the character of the linguistic expression "I" is a rule that always picks out the speaker in the context of utterance. Character is thus stable across contexts. By contrast, content is context-sensitive: the content of an indexical is that which it refers to in a given context. When I utter the expression "I" in a given context, it refers to me, while when you utter the same expression in a given context, it refers to you.

[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.

While perceptual representations differ from linguistic representations, we can use Kaplan (1989)'s framework to help explicate the *Dual Contents View*. The relevant contexts are no longer contexts of utterance, but contexts of use: the perceptual situation in which an egocentric reference frame is used with a perceptual representation. Given this, the view takes the character of an egocentric reference frame to be a rule that picks out a pair – a location and an individual – in the context of use. The content of such a frame in a particular context of use is thus a specific location and individual. To see this more clearly, let's look at the cyclopean reference frame. We can express its character as follows:

(2) ⟦**CRF**⟧: worlds w and contexts $c \rightarrow \langle$location l, individual $i\rangle$ in c, where l is occupied by the midpoint between the eyes of i.

Like the linguistic expression "I," the cyclopean reference frame has a stable character and variable content. The frame's character is a rule that always picks out a location and an individual in every context. Its content in a context is a particular location and individual. When I use the frame with my perceptual representation of my computer, the frame refers to a location (that occupied by the midpoint between my eyes) and my self. When you use this frame with your perceptual representation of your computer, the frame refers to a different location (that occupied by the midpoint between your eyes) and yourself.

The *Dual Contents View* takes egocentric reference frames to partially resemble individual indexicals and partially resemble locational indexicals. Like the linguistic expression "I" or the conceptual self-representation used in thought, part of the character of egocentric reference frames maps contexts to individuals. Thus, part of their content in a given context is an individual, just as it is for "I." At the same time, part of the character of egocentric reference frames maps contexts to locations, like the linguistic expression "here" or its conceptual analogue. As such, egocentric reference frames and "here" also have similar contents: both refer to a location in a context.

Of course, what makes egocentric reference frames distinctive according to the view is that they feature two kinds of indexicality simultaneously. Unlike individual and locational indexicals in thought and language, egocentric reference frames pick out *both* an individual and a location in a context. Thus, we might think of the content of an egocentric reference frame as roughly analogous to saying (or thinking) "I am here." Nonetheless, an important distinction between the indexicals remains: while one can use "I" and "here" independently of one another in thought and language, egocentric reference frames package these two indexical components together in perception. The cyclopean reference frame, for instance, does not refer to just a location or just an individual; it always refers to both together.

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 two indexical components of egocentric reference frames are related by a constraint. This constraint, which varies across different frames, specifies which point on the relevant individual's body occupies the relevant location. For the cyclopean reference frame, the location referred to must be occupied by the midpoint between the eyes of the individual referred to.[4] For the head-centered reference frame, the relevant location must occupy the center of the relevant individual's head. Notice that if an individual moves their body such that the center of their head occupies the location formerly occupied by the midpoint between their eyes, their head-centered reference frame will pick out the same location and individual as their cyclopean reference frame did previously.

[4]I assume that context determines the contents of the two indexical components independently, and that these contents must then satisfy the constraint linking them. However, there are other ways to construe the function from contexts to contents. For example, one could take the location to be determined first, such that the relevant individual would be whoever was such that the midpoint between their eyes occupied the location referred to. This would seem to treat the cyclopean reference frame's spatial function as primary. Alternatively, one could take the individual to be determined first, such that the relevant location would be whichever the midpoint between the individual's eyes occupied. This, by contrast, would seem to treat the cyclopean reference frame's perspectival function as primary.

The frames are nonetheless distinguished by the constraints they satisfy: initially, the cyclopean reference frame's constraint is satisfied by the fact that the relevant location is occupied by the midpoint between the individual's eyes, while the head-centered reference frame's constraint is subsequently satisfied by the fact that the relevant location is occupied by the center of their head.

As discussed, the accuracy conditions of perceptual representations suggest that egocentric reference frames have dual contents. Further support for this view comes from reflecting on the function of such frames. Egocentric reference frames seem to perform two functions. First, they act as spatial reference frames: they are used to specify spatial relations relative to an origin point. Second, they act as egocentric frameworks: they specify relations relative to the perceiver and so capture the perceiver's point of view. The *Dual Contents View* explains these two functions of egocentric reference frames as follows: indexing to a location establishes the origin of a spatial reference frame, while indexing to an individual specifies whose perspective the origin represents. Both contents are necessary for egocentric spatial representation. Without indexical individual contents, indexical locational contents could establish the origin of an allocentric reference frame, while without indexical locational contents, indexical individual contents could establish a non-spatial egocentric framework.

The idea that egocentric reference frames perform dual functions parallels Burge (2011)'s analysis of what he calls an egocentric index. According to Burge, an egocentric index is constitutive of perceptual representational content (p. 293). While his account of egocentric indices differs from the view I have presented of egocentric reference frames, Burge similarly attributes two functions to the former.[5] One function is "to index an origin for a framework of representation" (p.

[5] Burge's account of what egocentric indexes refer to is somewhat ambiguous. In "Self and Self-Understanding" (2011), he writes that they refer to "the individual – or, more commonly, some position on the individual or some time at which the perception occurs" and that they "single out an individual, position, or time in a context-dependent way" (p. 294). Here, it seems somewhat unclear whether reference to individuals, positions, and times is mutually exclusive or whether indexes might refer to combinations of these elements. In later work, "Psychological Contents and Egocentric Indexes" (2019), Burge seems to suggest that these kinds of reference are mutually exclusive. He writes that "egocentric reference to individuals probably emerges only when an individual can represent its own body and coordinate such representation intermodally with spatial and temporal egocentric indexes" 2019, p. 50). Not only does reference to individuals emerge later, but it is merely "coordinated" with indexes which refer to positions or times. On this reading, his view differs from mine. According to the *Dual Contents View*, a single egocentric reference

294). In other words, 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). That egocentric indices have ego-relevance seems closely related to the idea that egocentric reference frames capture an individual's point of view.

I have offered an analysis of egocentric reference frames according to which they carry indexical locational contents and indexical individual contents. As discussed, this *Dual Contents View* comports with an account of the accuracy conditions of perceptual representations, which takes them to include both the perceiver and her location. Moreover, this view explains how egocentric reference frames fulfill two distinct functions insofar as they serve as spatial reference frames and capture a perceiver's point of view.

2.3 Two arguments for implicitness

Having offered a view of the contents of egocentric reference frames, I now turn to consider how they are used with perceptual representations. Specifically, are egocentric reference frames constituents of complex perceptual representations or do they remain implicit in such representations? This question closely parallels an existing debate in the literature that concerns whether indexical individual contents are part of the contents of perceptual representations or rather remain implicit (Bermúdez, 1998; Campbell, 1994; Ismael, 2012; Musholt, 2015; Recanati, 2007, 2012; Schwenkler, 2014). Framing the question in terms of egocentric reference frames helps anchor it to the empirical study of perception, making it more directly informed by – and answerable to – the empirical literature. At the same time, the philosophical debate about whether the self is implicit in perceptual content provides arguments and considerations that help motivate my answer to the question about egocentric reference frames. In what follows, I will argue that egocentric reference

frame refers to both an individual and a location. Although I will not pursue the possibility here, I note that this kind of frame may also refer to a time.

frames remain implicit and present two arguments for this view: one based on considerations of simplicity and another based on considerations of the kinds of errors that perceptual systems make.

To make our question about the role of egocentric reference frames in perceptual representations more concrete, let us return to our previous example in which R is a complex perceptual representation used with the cyclopean reference frame. I suggested that the object o, distance d, and direction θ appear in R's accuracy conditions in virtue of constituent representations of an object, distance, and direction. I then proposed that the location l and midpoint between the eyes of individual i figure in R's accuracy conditions in virtue of the use of the cyclopean reference frame. But does this mean that R includes a constituent representation of the cyclopean reference frame?

There are two ways to answer this question. According to what I will call the *Explicit Model*, perceptual representations include constituent representations of the egocentric reference frame relative to which the locations of objects are specified. Thus, this view takes reference frames to be represented in the same way as other elements of the perceptual scene: just as complex perceptual representations include constituent representations of objects, distances, and directions, they include constituent representations of egocentric reference frames. Let **o** be an object representation, **d** be a distance representation, θ be an angle representation, and **CRF** be a cyclopean reference frame representation. According to the *Explicit Model*, complex perceptual representations are structured as follows:

(3) $\langle \mathbf{o}, \mathbf{d}, \theta, \mathbf{CRF} \rangle$

What I call the *Implicit Model* offers an alternative. It takes perceptual representations to include constituent representations of only an object, distance, and angle and not a constituent representation of the relevant egocentric reference frame. Thus, according to the *Implicit Model*, complex perceptual representations are structured as follows:

(4) $\langle \mathbf{o}, \mathbf{d}, \theta \rangle$

Critically, the *Implicit* and *Explicit Models* agree on the accuracy conditions of perceptual representations. Both hold, for example, that R is accurate if and only if the conditions in (1) obtain. With respect to *R*, they differ only on whether it includes a constituent representation of the cyclopean reference frame. According to the *Implicit Model*, *R* does not include such a constituent; rather, the location *l* and the midpoint between the eyes of individual *i* figure in its accuracy conditions without being explicitly represented. In this sense, the cyclopean reference frame remains implicit. The *Implicit Model* maintains that this is true more broadly: egocentric reference frames always remain implicit in perceptual representations.

2.3.1 Simplicity

The first argument for the *Implicit Model* draws on a philosophical strategy famously employed by Perry (1986). Perry imagines a case in which indexical locational contents might remain implicit – or as he says, "unarticulated" – in conceptual thought. Specifically, he asks us to consider Z-landers, a group of people who have only ever lived in Z-land, only communicate with fellow Z-landers, and do not know of other lands. As such, Z-land plays an invariant role in Z-landers' location-sensitive thoughts: all such thoughts concern Z-land and only Z-land. Given this, Perry explains that "there is no need to postulate a concept or idea of Z-land as a component of their thought in order to secure the connection to Z-land" (p. 144). Doing so would be unnecessary. Instead, it is simpler to posit that Z-land is an "unarticulated constituent" of Z-landers' location-sensitive representations.

Proponents of the view that indexical individual contents are implicit in perceptual representations often appeal to Perry's style of argument for support (Ismael, 2012; Recanati, 2012).[6] For example, Recanati uses it to analyze proprioception. He first notes that proprioceptive information is invariant insofar as it always and only concerns oneself. He then argues that because "the role of

[6]Ismael (2012) contrasts representational contents in experience with those in language and thought. She notes that representational contents in language and thought "have an argument place for the subject that can take a range of values and that acts as a separate degree of representational freedom" (p. 72). She then explains that we lack the capacity to ascribe properties to different subjects in experience. It is invariant that our experiences are our own. Therefore, Ismael concludes that there is no argument place for subjects in the representational contents of experiences: "To experience pain is to be aware of a certain state, but there is no internal dimension of variation in that content of that state that corresponds to different values for the subject" (p. 72).

representational content is to keep track of whatever is susceptible to variation," there is no need for the subject to appear in the contents of her proprioceptive representations (p. 189). Positing this would be superfluous. It is simpler to assume that the self does not figure in the contents of perceptual representations.

We can apply this reasoning to the use of egocentric reference frames in perception. Just as the self's role in proprioception appears invariant, so too is the use of egocentric reference frames in perception. More specifically, it seems that a particular egocentric reference frame is used invariantly at a particular stage of perceptual processing. Support for this comes from the observation that different egocentric reference frames appear to be implemented in different brain regions (see Andersen et al., 1985; Barendregt et al., 2015; Gross & Graziano, 1995; O'Keefe & 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 to be fixed. That is, perceived objects are represented at locations first relative to one egocentric reference frame, and then a second, third, and so on. For example, in visual perception, objects are initially represented relative to a retinotopic reference frame; as processing continues, they come to be represented relative to the cyclopean and then the head-centered reference frames (Barendregt et al., 2015; Grossberg et al., 1993). Given this invariance, considerations of simplicity suggest that we should not posit explicit representations of such frames. Doing so would be unnecessary. Thus, we have reason to choose the *Implicit Model*.

Beyond theoretical parsimony, there may be computational reasons to prefer the simpler model. This is because the simpler model will describe a system that is less cognitively demanding to operate. This consideration further supports the *Implicit Model*: by avoiding the need to explicitly represent egocentric reference frames, our perceptual systems may conserve time and computational resources.

It is important to recognize that this argument from simplicity provides only *prima facie* support for the *Implicit Model*. Simplicity considerations apply when competing models are equivalent in all other respects. One might wonder whether explicitly representing egocentric reference frames could support additional computational functions that would render our two models functionally non-equivalent. The burden of proof falls on defenders of the *Explicit Model* to demonstrate what explanatory work representations of egocentric reference frames might accomplish. In the meantime, there is a stronger argument for the *Implicit Model* that sidesteps this issue of functional equivalence.

2.3.2 Empirical fit

I now turn to argue that the *Implicit Model* better describes the kinds of representations that perceptual systems actually use. If egocentric reference frames were constituents of perceptual representations – i.e. if the *Explicit Model* accurately described the role of such frames in perceptual representations – we should expect to find evidence of two specific kinds of representations involving these frames that we don't, in fact, find. As such, the *Implicit Model* seems to better match the perceptual evidence.

Beyond perceptual representations that specify the location of an object relative to an egocentric reference frame (as shown in 3), the *Explicit Model* allows for two other kinds of representations. First, it allows for what I call *circular representations*. An example of a circular representation is shown below:

(5) $\langle \mathbf{CRF}, \mathbf{d}, \theta, \mathbf{CRF} \rangle$

Complex representations of this form use a constituent representation of the same egocentric reference frame twice. In so doing, they represent that frame as located at some distance *d* and angle θ away from itself. For any distance and angle greater than zero, this generates semantic incoherence. The reference frame problematically serves as both the coordinate system relative to which the locations of objects are specified and as the object whose location is specified. One might compare this to the English sentence "Here is five feet away from here." As in (5), the dual use of the same locational indexical makes this sentence uninterpretable.

Second, the *Explicit Model* allows for *allocentric representations*. As I use the term here, these are complex representations that specify the location of one object relative to another (which is external to the perceiver). In such representations, a constituent representation of the second object replaces the constituent egocentric reference frame representation, as shown below:

(6) $\langle \mathbf{o_1}, \mathbf{d}, \theta, \mathbf{o_2} \rangle$

I argue that neither circular nor allocentric representations are used in early to mid-level perceptual processing. Two sources of evidence support this claim: reflection on perceptual experience and neural findings.

Perceptual experience provides reason to doubt that circular representations are used in perception. Because circular representations are semantically incoherent, it is difficult to say what kind of experience their use would correspond to. Surely, they do not account for typical perceptual experiences. Perhaps the most plausible option for an atypical experience they might correspond to is heautoscopy, in which individuals perceive a double of themselves and experience an ambiguous sense of first-person perspective and self-location (Blanke & Metzinger, 2009, p. 10). In some cases, patients report feeling as though they occupy both their physical body and the observed double simultaneously. For example, one woman reported experiencing herself lying in bed and simultaneously standing at the foot of the bed; when asked where she took herself to be, she answered, "I am at both positions at the same time" (Blanke et al., 2004, p. 248). It might seem plausible that in this case, the woman represents one egocentric reference at two different locations, which correspond to the two locations she takes herself to be. However, this representational account of such experiences is likely mistaken. Heautoscopic experiences are generally associated with vestibular hallucinations and changes to the body schema in the temporoparietal junction (Blanke & Mohr, 2005, p. 92), pointing to processing errors well beyond those that might be found in the basic spatial representations used by individual perceptual systems.

Phenomenological evidence also suggests that we never use allocentric representations in perception. Of course, standard perceptual experiences are always experienced from one's own point of view. But even when perceptual experiences deviate from the norm, they generally remain experiences from a point of view that one experiences as one's own. Consider out-of-body experiences, in which people feel like they are located outside of their physical body. In such cases, one's point of view is likewise experienced as located outside the body (Blanke & Metzinger, 2009, p. 9). Critically, however, one still experiences this point of view as one's own (Blanke & Metzinger, 2009, p. 9). Thus, such experiences remain importantly different from our experiences of allocentric representations like maps, which fail to include the perceiver's point of view. This provides some evidence for the claim that even in unusual circumstances, perceptual systems use only egocentric representations.

Let's now turn to the neural evidence. There, we find an absence of evidence for the use of circular representations. To the best of my knowledge, there is no evidence suggesting that perceptual representations ever include two constituent representations of one and the same egocentric reference frame. The kinds of errors known to occur are those that involve the misrepresentation of an object, its properties, or its location. Nor is there neural evidence for the use of allocentric representations in early or mid-level perceptual processing. Early visual areas are primarily represented relative to either a retinotopic or cyclopean reference frame (Barendregt et al., 2015). Where we find allocentric representations is in the hippocampus and related medial temporal lobe structures, which support navigation and spatial memory (Epstein et al., 2017; Ekstrom et al., 2017; O'Keefe & Dostrovsky, 1971. While these allocentric representations may influence perception through top-down modulation, there remains no evidence that early visual areas themselves token allocentric representations.

The absence of circular and allocentric representations in early perceptual processing poses a serious challenge for the *Explicit Model*. By taking perceptual representations to include explicit representations of egocentric reference frames, the *Explicit Model* allows that perceptual systems could, in principle, token either of these kinds of representations. Yet neither phenomenological nor neural evidence supports their use in early perceptual processing. Because it fails to fit the empirical facts, the *Explicit Model* thus appears ad hoc: it leaves unexplained why only some of the representations it permits are used, while others are not.

By contrast, the *Implicit Model* is better tailored to the data. Because it does not treat egocentric reference frames as explicit constituents of perceptual representations, the model rules out the possibility of circular or allocentric representations. As a result, the representations it posits better match the representations that perceptual systems seem to use. Moreover, the *Implicit Model* can explain why circular and allocentric representations are never tokened: perceptual systems cannot token them, and so they do not token them.

One might suggest that the *Explicit Model* could include rules that prohibit the production of circular and allocentric representations. If such rules were included, then the *Implicit* and *Explicit Models* would equally well match and explain egocentric perceptual experiences. But this suggestion is unappealing for two reasons. First, adding rules to the *Explicit Model* would only increase its relative complexity and thereby strengthen the simplicity considerations in favor of the *Implicit Model* (§2.3.1). The former model would posit not only a more complex representational structure – which would include constituent representations of egocentric reference frames – but also additional rules constraining how those representations could be used. Both posits are unnecessary. The *Implicit Model* naturally excludes circular and allocentric representations in virtue of its simpler representational structure, obviating the need for additional rules.

Adding rules to the *Explicit Model* is misguided for a deeper reason. Representations play a certain role within the mind: they carry information whose use is variable. As theorists, we should posit representations only when information is used variably. If the use of information is fixed, then it should not be described as represented. The proposed change to the *Explicit Model* would fix the use of egocentric reference frames in perceptual processing, making it a mistake to describe such frames as represented. In short, I take it to be a kind of theoretical error to hold both that egocentric reference frames are represented and that their use is fixed.

The question of whether egocentric reference frames are explicitly represented in perception connects to a broader philosophical discussion about whether the self is explicitly represented in perception. Musholt (2015) argues that it is not and offers a structurally similar argument to the one I have given here. She first notes that if the self were explicitly represented, "there should, at least in principle, be the possibility of misrepresentation" (p. 81). But experience does not seem to include any such misrepresentations: I do not have pains that I ascribe to someone other than myself, nor do I have visual experiences that I ascribe to someone other than myself. Rather, our experiences always correctly identify our selves as their subject. She takes this to support the conclusion that the self is not explicitly represented in perception.

While Musholt (2015) provides valuable groundwork, focusing on egocentric reference frames offers a more empirically tractable approach to determining whether indexical individual contents are explicitly represented in perception. Rather than searching for general misrepresentations of the self, I have looked for two specific types of representational errors: circular representations that would locate reference frames relative to themselves, and allocentric representations that would abandon the first-person perspective entirely. The absence of these specific error types provides strong empirical evidence for the conclusion that egocentric reference frames – and thus, indexical individual contents – remain implicit in perception.

2.4 Architectural encoding

If, as I argued in the previous section, egocentric reference frames are not constituents of perceptual representations, then how do they nonetheless contribute to the accuracy conditions of such representations? Perhaps, the information they carry is located elsewhere in the mind. In this section, I will argue that egocentric reference frames are encoded in the functional architecture of one's perceptual systems. My argument relies on a particular view of functional architecture, according to which it does not merely compute over representations but also encodes information. After explicating this view in greater detail, I will show how this framework can be used to understand the role of egocentric reference frames in perception.

2.4.1 Architecturally encoded information

The functional architecture of a representational system enables the system to process representations. It makes possible algorithms or functions that take certain representations as inputs and produce certain representations as outputs. I want to propose that the functional architecture of a representational system itself encodes information. According to this view, information is encoded in two places within a representational system: in the representations themselves and in the functional architecture that processes those representations.

This view, while not widely adopted, appears across different literatures. Musholt (2015), concerned with how information about the self becomes explicit, draws on developmental psychologist Karmiloff-Smith's theory of representational redescription. Musholt writes that at the lowest stage of this process, "information is encoded in procedural form – in procedures for processing and reacting to environmental inputs..." (p. 103). Her emphasis on procedures suggests that she is concerned with what I call the functional architecture of a representational system. She clarifies what it means for information to be encoded when she adds: "The information is *in* the system, but not available *to* the system" (p. 103). This points to an important contrast between functional architecture and representations that I will return to later. The idea that functional architecture encodes information also appears in Johnson (2020)'s account of implicit bias. She writes that "certain stereotype-like contents can also be encoded by embedding them in non-representational transformation rules" (p. 1203). For Johnson, architectural transformation rules provide an alternative mechanism for encoding stereotype-like contents beyond explicit representations.

Let's explore the idea that functional architecture encodes information using a visual example. A particular shading pattern on the retinal image is compatible with various environmental causes: it can be caused by light from above hitting a concave shape or by light from below hitting a convex shape. The visual system resolves this ambiguity in a way that is consistent with the assumption that light comes from above, producing a representation of a concave shape (Palmer, 1999, pp. 244-245). According to the proposed view, the system does not merely function consistently with this assumption. Rather, the system's architecture itself encodes the information that light comes from above.

Positing that functional architecture encodes information has two advantages. The first is explanatory. In the example above, ascribing this information to the system's architecture helps explain how the system comes to represent a concave shape: it does so by processing a particular shading pattern using the information that light comes from above. This offers a better explanation of the system's functioning than the orthodox view, according to which this information merely describes the system's functioning. Tyler Burge expresses this standard view when he maintains that "principles governing transformations in perceptual systems are not the contents of any perceptual state" but rather "specify patterns that occur in perceptual systems" (p. 406). While he considers these principles "psychologically real," Burge treats them as external descriptions of perceptual patterns rather than as information encoded within the system's architecture. This creates an explanatory gap: if the system doesn't contain the information that light comes from above, we lack an explanation of why it seems to function in line with this constraint.

The second advantage concerns theoretical consistency. If we are willing to attribute information to cognitive systems at all, why restrict this attribution to explicit representations? Both explicit representations and functional architecture play essential roles in explaining cognitive functions. In our previous example, representations of the shading pattern together with the system's built-in assumption that light comes from above help explain why the system produces a representation of a concave shape. Given that representations and functional architecture are both critical to this explanation, we have reason to treat both as carrying information.

Treating both explicit representations and functional architecture as carrying information does not collapse the distinction between them. They are distinguished by how such information is used: representational information is used variably to track changing features of a context, whereas architecturally encoded information is fixed and constrains processing in the same way across contexts. For instance, in veridical cases, a visual representation of a concave shape is used only when a concave shape is present in the environment. By contrast, the information that light comes from above is always used, regardless of environmental lighting conditions. As theorists, we can thus look to how information is used to determine whether such information is explicitly represented or architecturally encoded.

2.4.2 Architecturally encoded egocentric reference frames

This view of functional architecture can be used to elucidate the role of egocentric reference frames in perception. Recall our question: how do a particular location and individual appear in the accuracy conditions of perceptual representations if such representations do not include an explicit constituent representation of an egocentric reference frame? The answer I want to suggest hinges on the idea that information about the location and individual is encoded in the functional architecture of perceptual our systems. In other words, egocentric reference frames are architecturally encoded in perceptual systems.

We can see how this architectural encoding of egocentric reference frames might work in practice by looking at visual processing. As discussed, different reference frames operate at different stages of visual processing. I propose that these different frames are architecturally encoded in the visual system. Early visual processing architecturally encodes a retinotopic reference frame. More specifically, it encodes indexical locational information and indexical individual information, such that the retina of the individual referred to occupies the location referred to. The next stage of visual processing encodes the cyclopean reference frame. This is to say that it encodes indexical locational and individual information, such that the midpoint between the eyes of the individual referred to occupies the location referred to. Later visual areas encode other egocentric reference frames in similar ways. This analysis does not merely describe how the visual system functions to represent space. Rather, it also claims that these bundles of indexical information are *in* the system's functional architecture.

We can also consider where in the brain these reference frames might be architecturally encoded. Given that V1's function is to map objects relative to a retinotopic frame, it would presumably encode the retinotopic frame (Barendregt et al., 2015). Similarly, given that V2's function is to map objects relative to the cyclopean reference frame, it would presumably encode the cyclopean reference frame (Barendregt et al., 2015). Later visual areas likely encode the reference frames used in later visual processing. While speculative, this suggests how these bundles of indexical information might be physically instantiated.

Interpreting egocentric reference frames as architecturally encoded naturally follows the advice of Pylyshyn (1984) on how to develop cognitive models. 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). 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.

This architectural encoding in the visual system parallels our light-from-above case. Just as the visual system's architecture encodes the information that light comes from above, it also encodes various egocentric reference frames. Critically, we can distinguish both kinds of architecturally encoded information from the information that the system explicitly represents. Only the architecturally encoded information is used invariantly across contexts: the system always relies on the light-from-above constraint and always employs the same egocentric reference frames at specific stages of processing.

We can now see how the architectural encoding of egocentric reference frames helps explain the accuracy conditions of perceptual representations. Since functional architecture, like explicit representations, carries information, it provides an additional mechanism for fixing accuracy conditions. Thus, a perceptual representation's accuracy conditions include a specific location and individual, not because these are explicitly represented, but because information about them is encoded in the perceptual system's functional architecture. This shows how egocentric reference frames can shape the accuracy conditions of perceptual representations without being explicitly represented.

Consider again representation *R*, whose accuracy conditions include a location *l* and an individual *i*, such that *l* is the midpoint between the eyes of *i*. The location and individual appear in these accuracy conditions precisely because information about them is architecturally encoded in the perceiver's visual system – because the cyclopean reference frame is architecturally encoded. In this way, *R* has these accuracy conditions without containing any explicit representation of the cyclopean reference frame.

I have proposed that functional architecture can itself encode information, thereby offering a mechanism besides explicit representations for fixing accuracy conditions. More specifically, I have argued that given their invariant use across contexts, egocentric reference frames are architecturally encoded in perceptual systems. This architectural encoding explains how specific locations and individuals can appear in the accuracy conditions of perceptual representations without being explicitly represented, thereby offering a mechanism by which egocentric reference frames shape perceptual accuracy conditions.

2.5 The encoded self

To sum up, I have proposed a view of egocentric reference frames according to which they carry two kinds of indexical information – locational information and individual information – and are architecturally encoded in our perceptual systems. The upshot, given my focus on the self, is that indexical individual information is baked into the hardware of our perceptual systems.

This aligns with certain reflections on how the self figures in perception. Consider Wittgenstein (1922), for example. In the *Tractatus*, Wittgenstein writes that "the subject does not belong to the world but it is a limit of the world" (5.632). This is followed by the thought that "from nothing in the field of sight can it be concluded that it is seen from an eye" (5.633). We can interpret these ideas as follows: just as an eye does not see itself in the visual field, so too, the subject does not find itself in the world. In representational terms, the eye and the subject can represent the visual field and the world, respectively, but cannot represent themselves. My account of egocentric reference frames offers a parallel: indexical individual information about the perceiver is embedded in the architecture of perceptual systems and constrains how they operate, yet it is not among the explicit contents of perception. In this respect, indexical individual information functions as a kind of "limit" on perceptual representations, rather than as an object of such representations.

We can also explain Hume's (1739/1978) observation that whenever he introspects, he "fails to catch" himself and "never can observe anything but the perception" (Bk. 1, ch. 4, §3). He can never catch himself because his indexical individual information is encoded in the functional architecture of his perceptual systems rather than explicitly represented. This information remains inside the walls of his mind, so to speak, and is therefore unavailable for conscious reflection. What he can observe – what is consciously accessible – are those things that are explicitly represented, what he calls "the perception." This line of thought allows us to explain Hume's experience without endorsing the conclusion that perception fails to involve indexical individual information – a conclusion that could be consistent with his rejection of the existence of the self.

While thus doing justice to the phenomenology of perceptual experience, the view I have presented also raises a new question concerning the relation between self-related information in perception and thought. According to the view I have developed here, indexical individual contents are architecturally encoded in our perceptual systems, but in thought, such contents are explicitly represented. The question, then, is what the representational pathway is from the former to the latter. In other words, how do architecturally encoded indexical individual contents come to be explicitly represented in thought? The next chapter begins to sketch an answer to this question.

Chapter 3

The Hierarchy of Selves in Perception[1]

In the previous chapter, I argued that indexical individual information is architecturally encoded in our perceptual systems. While my interest broadly lies in how this information transitions from being architecturally encoded in perception to being explicitly represented in thought, this chapter focuses on a narrower question: whether such information is ever explicitly represented within perceptual processing itself. A prevalent philosophical view holds that it is not. According to this view, indexical individual information always remains implicit in perceptual representation (Campbell, 1994; Ismael, 2012; Musholt, 2015; Recanati, 2007, 2012).[2] This reflects something like Wittgensteinian insight already considered. 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

[1]A revised version of this chapter is forthcoming in *Review of Philosophy and Psychology*.

[2]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."

representation.

I will argue against this view that indexical individual information is *always* implicit in perception. More specifically, I will argue that whether such information is architecturally encoded or explicitly represented 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 indexical individual information 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 §3.1, I articulate the view that indexical individual information is always implicit in perceptual representations. I discuss how this view follows from my account of egocentric reference frames if such frames are *always* architecturally encoded. But, this does not seem to be the case, as I argue in §3.2. Rather, I propose that implicit indexical individual information becomes explicitly represented as perceptual processing advances and one egocentric reference 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. §3.3 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*. In §3.4, I discuss how this hierarchy of egocentric reference frames partially explains one's capacity for complex movements and informs self-locating judgments in creatures with conceptual capacities. I briefly conclude in §3.5.

3.1 *Implicit Self*

As mentioned, a prevalent view in the literature holds that indexical individual information is implicit in perceptual representation (Campbell, 1994; Ismael, 2012; Musholt, 2015; Recanati, 2007, 2012). To better understand this idea, 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. 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. Call this view *Implicit Self*.

> ***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 myself.[3] 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

[3]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.

constituent representations of self, although the self figures in their accuracy conditions.[4]

My interest here is in how indexical individual information 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 indexical individual information 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 (see §2.3.1). 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

[4]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.

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 at which an error of *mis*identification could get introduced" (Morgan & 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.[5]

While these motivations make *Implicit Self* theoretically attractive, we should be cautious about accepting it as a general principle. This skepticism arises when we interpret how the view might work using egocentric reference frames. As we saw in the previous chapter, when an egocentric reference frame is used in its customary role as the reference frame relative to which objects are represented, it is architecturally encoded. Granting that egocentric reference frames carry indexical individual information, this entails that indexical individual information remains implicit in perceptual representation. The difficulty is that perception also involves specifying the locations

[5]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).)

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. In the next section, I argue that in such cases, we have reason to think that egocentric reference frames are explicitly represented and thus, that indexical individual information is explicitly represented. This suggests that contra *Implicit Self*, we have reason to posit a restricted use of representations of self in perception.

3.2 The *Nested Frames View*

The reason we should be cautious about taking *Implicit Self* to describe all perceptual representations is that egocentric reference frames are not limited to functioning as reference frames, as they do in standard cases; they also 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 argue, the view shows how the indexical individual information 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 & Zipser, 1988; Barendregt et al., 2015; Chen et al., 2013; Colby & Duhamel, 1996; Groh, 2014; O'Keefe & 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.[6] The

[6]The *Nested Frames View* seems to capture the representational relations between the egocentric reference frames

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.[7]

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 implicit and explicit roles. 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

used in human visual perception, as will be seen in §3.3. 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.

[7] 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.

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

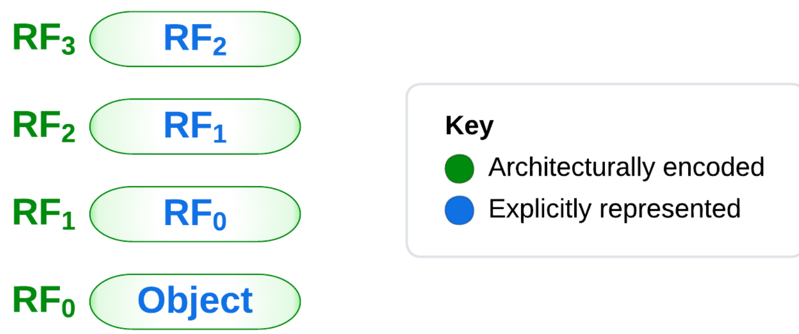

Figure 1: Nested Reference Frames. Perceived objects are nested within RF_0, RF_0 is nested within RF_1, RF_1 is nested within RF_2, and RF_2 is nested within RF_3.

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.[8] 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.[9] 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

[8]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 & Lacquaniti, 1994) and idiosyncratic reference frames (Chang & Snyder, 2010).

[9]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.

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 not only reinforce the general idea of a hierarchy of egocentric reference frames in perception but also specifically align with the structure outlined above 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 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 Figure 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).[10] 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

[10] A similar 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).

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 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 such 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 §3.3.1. 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 & 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 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, there is no need for it to be represented (§2.4.2). 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 indexical individual information 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 indexical individual information (§2.2).

Another source of support for the idea that explicit representation of an egocentric reference frame entails explicit representation of indexical individual information 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, indexical individual information – here, information that the eyes are *my* eyes – is also explicitly represented. More generally, this suggests that indexical individual information 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 indexical individual information 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 indexical individual information 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.

3.3 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 indexical individual information comes to be explicitly represented in perception.

3.3.1 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 Figure 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.

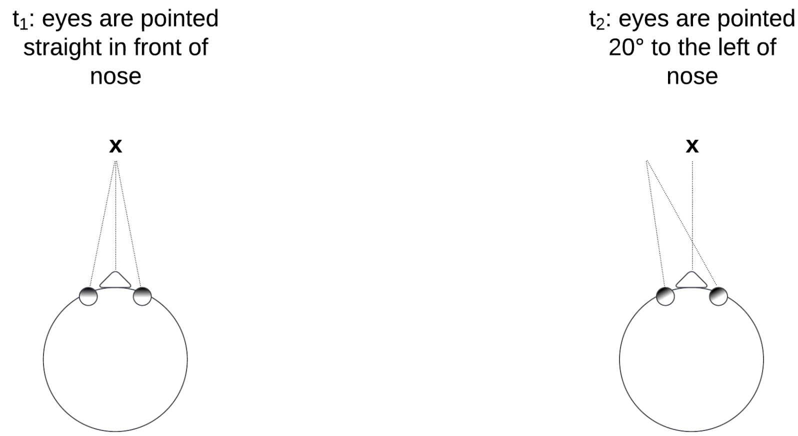

Figure 2: Gaze Shifts

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 Figure 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.[11]

[11] 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.

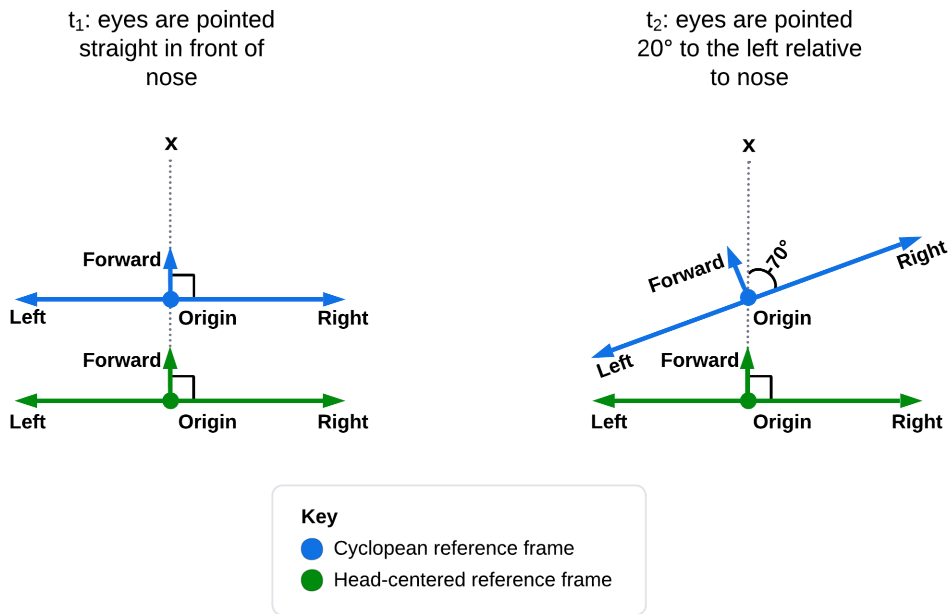

Figure 3: A Representational Account of Gaze Shifts

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 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 refer-

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.

ence 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° 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 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.[12]

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 & Miall (2008)'s results indicate just this.

In Balslev & 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.

[12]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).

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).[13] 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 & 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 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.[14] 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 located relative to one's head-centered frame, then we can conclude that gaze shifts involve the use of an

[13]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).

[14]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).

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 an architecturally encoded cyclopean reference frame and explicitly representing the cyclopean frame at a position relative to an 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*, indexical individual information is explicitly represented in perception. We have reason to believe that the cyclopean frame carries indexical individual information insofar as it seems to capture the individual's perspective (§2.2). When the frame is made explicit, so too is the information that it carries.

3.3.2 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 Figure 4, 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 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.

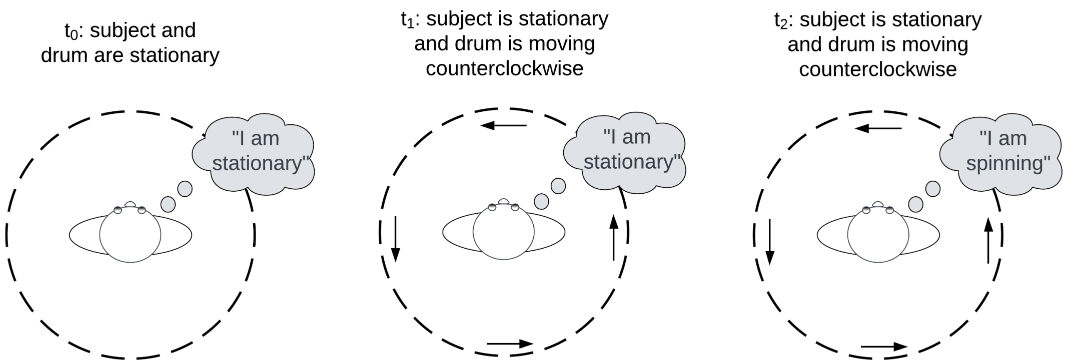

Figure 4: Vection

Vection is often understood as involving a change from one *rest frame* – the reference frame one takes to be stationary – to another (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. 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 Figure 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

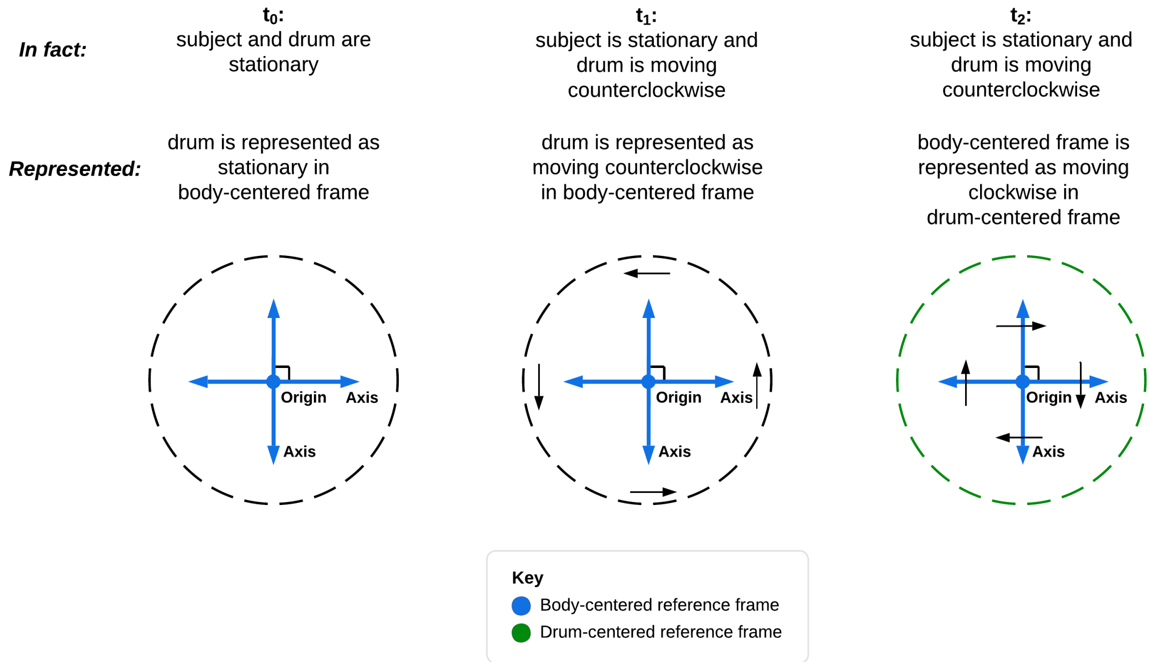

Figure 5: A Representational Account of Vection

frame, as shown in Figure 5.[15]

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 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. A natural way to encode the body as moving is to explic-

[15] 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.

itly represent it at different positions over time. 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 indexical individual information is explicitly represented in perception.

3.4 Implications of the hierarchy beyond perception

Having explicated the *Nested Frames View* and discussed how it accounts for gaze shifts and vection, I would like to highlight two important consequences of the view that extend beyond perception. One consequence concerns how the hierarchy of reference frames relates to a creature's capacity for movement; the other concerns how reference frames in the hierarchy inform *de se* thought. I'll develop each of these ideas in turn.

To begin, I propose that only creatures with the capacity for movement use a hierarchy of egocentric reference frames. 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 that 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. This 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 use a hierarchy of such frames. Thus, where there is no capacity for movement, we find at most one reference frame used by a creature in perception.

Increasingly sophisticated forms of movement, both within and across creatures, suggest the use of increasingly nested reference frames for two interrelated reasons. First, as movements become more complex, they engage the use of more bodily joints. For example, turning one's head while moving one's eyes involves more joints (both the eye sockets and the neck) than either movement involves individually. The joints, in turn, introduce degrees of physical freedom that a creature must track in cognition (as discussed in §3.2). The greater the degrees of physical freedom introduced by the joints, the greater the number of reference frames a creature requires. So, increasingly sophisticated kinds of movement typically require increasingly nested structures.

Second, each nesting reference frame marks the capacity to represent proprioceptive information that can be used to guide a creature'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. This proprioceptive information is critical to successful action because it allows one to perceive one's actions while in progress and adjust them if and as necessary. So, increasingly sophisticated forms of movement seem to be partially explained by the use of more nested reference frames insofar as they mark the availability of proprioceptive information.

One consequence, then, of the hierarchy of egocentric reference frames is that it seems to help explain a creature's capacity for complex movements. I'll now explore a second consequence of the hierarchy, which concerns its connection to *de se* thought. I conjecture that, in creatures with conceptual capacities, representations used with frames towards the top of the hierarchy are more likely to inform conceptual judgments about self-location than representations used with frames towards the bottom. I'll gloss a self-locating judgment as a judgment that specifies the spatial relations between oneself and the greater environment. This kind of judgment is often informed by perception. For example, one might form the self-locating judgment *'Buckingham Palace is to my left'* based on one's current perceptual state. By analyzing how perceptual representations inform self-locating judgments like this one, we can start to understand how the hierarchy of egocentric reference frames used in perception connects to the representation of self used in thought.

Borrowing an example from Peacocke (1992), let's 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 the left in your head-centered frame. If prompted to judge where the palace is relative to you, my intuitions are that 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. Thus, it is the representation used with the higher frame in the hierarchy 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 Buckingham Palace falls in the middle of your head-centered frame, but off to the left in your body-centered frame. Again, you seem more likely to judge that the palace is to your left and again, your judgment is based on a representation used with the higher frame in the hierarchy: the body-centered frame rather than in the head-centered frame. Although they concern different reference frames, what both of these cases suggest is that representations used with less embedded frames are more likely to inform one's self-locating judgments than representations used with more embedded frames.

There is some empirical confirmation of this hypothesis. Alsmith et al. (2017) used what they call a "Misalignment Paradigm" to evaluate the relative contributions of one's head- and body-centered reference frames to self-locating judgments. 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 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 representations used with reference frames towards the top of the hierarchy.

To wrap up, I have explored two consequences of the hierarchy of reference frames posited by the *Nested Frames View* that extend beyond perception. First, the capacity to perform complex movements seems to be partially explained by the use of increasingly nested structures. Second, in creatures with conceptual capacities, representations used with reference frames at the top of the nested structure seem to inform one's self-locating judgments more than representations used with frames at the bottom. So, the more sophisticated one's movements, the more reference frames compose one's hierarchy. And the closer a frame is to the top of the hierarchy, the greater its impact on one's conceptual representation of self.

3.5 Conclusion

This chapter has offered a new account of how the self is represented in perception. I have proposed that the egocentric reference frames used in perception form a hierarchy. The hierarchy reveals how indexical individual information 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. What's more, I have discussed how the hierarchy of egocentric reference frames not only seems to partially explain one's capacity for complex movements but also seems to inform self-locating judgments in creatures with conceptual capacities. 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.

Chapter 4

Local and Global Body Ownership: A Case for Representational Independence

Thus far, I have examined how representations of self might be used in exteroception, in our perception of the external world. But plausibly, we also use representations of self in interoception, when we perceive our own bodies. This possibility is particularly salient with respect to the sense of bodily ownership. Many philosophers posit that a feeling of bodily ownership, sometimes referred to as a feeling of "mineness," is a real and distinctive aspect of our phenomenology (Billon, 2017; de Vignemont, 2018; Peacocke, 2014). It is thought to be qualitatively different from the feelings that one has towards external objects, including those that one materially possesses.

A growing body of empirical evidence suggests that the sense of bodily ownership comes in two forms: a sense of ownership of individual body parts and a sense of ownership of our whole body. This raises two representational questions: What are the representational bases of local and global ownership? And, are these representational bases one and the same?[1] Answering these questions offers to shed light on how self-representation operates in bodily perception and to further develop our understanding of self-representation in perception more broadly.

[1]De Vignemont (2022) and Orbán & Wong (2023) refer to versions of these questions as "the problem of body mereology" and the "mereological question," respectively.

Philosophical accounts of the sense of bodily ownership have primarily been concerned with the sense of ownership that we have for our body parts, leaving these broader representational questions largely unaddressed. One exception comes from Bermúdez (2017), who proposes that both local and global ownership are grounded in a hierarchy of body part representations centered on the bodily joints. In this way, he suggests that the two forms of bodily ownership share the same representational basis.

However, recent empirical evidence suggests that the representational bases of local and global ownership may differ. In this chapter, I argue that while Bermúdez (2017)'s hierarchical account seems to successfully explain local ownership, it cannot also account for global ownership. I propose instead that global ownership is grounded in a single representation of self. This representation is used with a representation of a perceived object to ascribe ownership of the object to the perceiver. As I discuss, this representation is distinct from those that compose the hierarchy of body part representations in two ways: it is a single representation and it lacks spatial content. In both respects, this representation resembles the conceptual representation of self used in thought.

I will begin by outlining empirical evidence for a sense of bodily ownership and explain how this evidence reveals bodily ownership to have two distinct objects: the individual body parts and the whole body (§4.1). I will then present Bermúdez (2017)'s account of bodily ownership and show how it successfully covers many cases of body part ownership (§4.2). Next, I will argue that it cannot account for whole body ownership, given empirical evidence suggesting a dissociation between the representations that ground body part ownership and those that ground whole body ownership (§4.3). I will then propose that unlike body part ownership, whole body ownership is grounded in a single representation of self (§4.4). After discussing how this representation contrasts with the representations that ground body part ownership, I will relate this global ownership representation to the conceptual representation of self used in *de se* thought (§4.5).

4.1 Evidence for the sense of bodily ownership

The sense of bodily ownership refers to the unique feeling that our bodies are our own. Not only do multiple lines of evidence converge to support that this feeling is genuinely part of our phenomenology, but also, recent research suggests that this feeling can be directed towards two different kinds of objects: the individual body parts and the whole body. This section first reviews evidence for a sense of body part ownership and then examines evidence for a sense of full body ownership. As I discuss, these different kinds of ownership raise a representational question: are both grounded in the same representations?

One of the most significant pieces of evidence that supports positing a sense of bodily ownership as an aspect of phenomenology is the *rubber hand illusion*. Following the original experimental design used by Botvinick & Cohen (1998), a rubber hand is aligned next to the subject's real hand, and a screen obscures their real hand from view. The experimenter uses two paintbrushes to stroke the subject's real hand and the rubber hand simultaneously. Following the experiment, subjects report that when both hands were stroked simultaneously, the rubber hand felt as though it belonged to them and felt like a part of their body. They also report feeling as though the touches they felt were caused by touches to the rubber hand. In other words, subjects seem to experience a sense of bodily ownership for the rubber hand.

But do subjects in these studies *really* feel as though the rubber hand is part of their body? Evidence from nonverbal tests suggests that they do.[2] For example, when the two hands are stroked simultaneously and the rubber hand is threatened, subjects exhibit fear responses: they flinch as if their real hand were threatened. Moreover, this behavioral response is accompanied by an increase in skin conductance (Armel & Ramachandran, 2003). Skin conductance response measures the

[2] Admittedly, some of the evidence is mixed. For example, Botvinick & Cohen (1998) found evidence of "proprioceptive drift" of the real hand toward the location of the rubber hand. That is, participants' proprioceptive sense of their real hand's location appeared displaced in the direction of the rubber hand. However, further research suggests that proprioceptive drift may not be caused by the illusion of ownership (Holmes et al., 2006). Another purported physiological correlate of ownership is skin temperature. Moseley et al. (2008) found that the temperature of the real hand decreased during the subjective experience of ownership for the rubber hand, but subsequent research suggests that this may not be a strong indicator of the sense of bodily ownership (Rohde et al., 2013).

electrical conductance of the skin and indicates arousal of the autonomic nervous system. When elevated, it is interpreted as evidence of a "fight or flight" response. Thus, elevated skin conductance in response to threats to the rubber hand suggests that subjects genuinely feel threatened when the rubber hand is threatened.[3]

A second source of evidence in favor of positing a sense of bodily ownership comes from somatoparaphrenia. Somatoparaphrenic patients, who often suffer damage to the right hemisphere of the brain, deny that certain body parts (usually their contralesional limbs) are their own. Such denials are accompanied by further "illusional, confabulatory or delusional ideas" (Gerstmann, 1942, p. 909). For example, patients often attribute ownership of their disowned limb to someone they know, such as a relative or caregiver. One woman discussed her own left hand as follows: "It's my niece's hand...She is so kind, she left it here to keep me company... My niece is so sweet... but she's also very absent minded... look here, she was rushing home and forgot her hand here!" (Gandola et al., 2012, p. 1177). A plausible explanation of reports like this is that they accurately reflect patients' feelings of bodily disownership. One way we might cache out a sense of disownership is to posit that it results from a disruption to the sense of bodily ownership. If this is correct, then somatoparaphrenia thus gives us additional evidence for a sense of bodily ownership.

The rubber hand illusion and somatoparaphrenia appear to reveal different kinds of disruptions to the sense of bodily ownership. In the former, individuals report feelings of ownership for something not part of their bodies, while in the latter, they report feelings of disownership for a body part. Other purported disruptions to the sense of bodily ownership include the tool ownership illusion, the disownership illusion, and phantom limb syndrome (see de Vignemont, 2018). Taken together, these cases all point to the same conclusion: there is a sense of bodily ownership, and it can be disrupted.

Importantly, these sources of evidence all point to the existence of a feeling of ownership for

[3] In line with these findings, Ehrsson et al. (2007) found that the strength of the illusion of bodily ownership seems to correlate with neural activity in areas of the brain associated with anxiety.

individual body parts. In both the rubber hand illusion and somatoparaphrenia, it is a feeling of ownership for one's hand that is manipulated or disrupted. I now turn to consider evidence that we also have a feeling of ownership for our whole bodies.

Researchers have modified the experimental paradigm used in the rubber hand illusion to test for a sense of ownership for the whole body. This is known as the *full-body illusion*. In one of the first studies of the full-body illusion by Petkova & Ehrsson (2008), participants looked down at their bodies while wearing head-mounted displays. The head-mounted display showed a live feed from a video camera centered on a mannequin's head, positioned so as to be looking down at its body, thus mirroring the bodily position of the participants. As in the rubber hand illusion, the experimenter used two paintbrushes to stroke the subject's abdomen and the mannequin's abdomen synchronously in the test condition and asynchronously in the control condition. Following the experiment, participants completed a questionnaire in which they were asked to agree or disagree with statements on a seven-point scale. Participants were more likely to agree with the statement "It felt like the mannequin's body was my body" following synchronous touching than following asynchronous touching. This difference was statistically significant.

Subjective reports of a sense of ownership for the mannequin's body were further confirmed by physiological evidence. During a second experiment, participants observed the experimenter's hand appearing to cut the mannequin's abdomen with a knife following one minute of either synchronous or asynchronous stroking. Participants' skin conductance responses were significantly greater following synchronous stroking. Petkova & Ehrsson (2008) suggested that "the participants' emotional systems responded as one would anticipate a person to respond were their own body threatened" (p. 3).

One might express skepticism about whether participants in this experiment really experienced a sense of ownership for the mannequin's whole body, rather than merely experiencing a sense of ownership for its abdomen. Perhaps participants agreed with the statement that the mannequin's body felt like their own because it was approximately true. What they really felt, however, was that the mannequin's abdomen felt like their own. According to this interpretation, the experiment manipulated the sense of body part ownership.

One piece of evidence against this skeptical interpretation is the observation of a "spread of ownership" from one body part to another. In a third experiment, Petkova & Ehrsson (2008) changed the location of the synchronous and asynchronous stroking from the abdomen to the hand. Then, as before, the experimenters threatened the mannequin's abdomen using a knife. Participants again exhibited elevated skin conductance response in the synchronous condition. In other words, synchronous stroking of the hand appeared to produce a feeling of ownership for the abdomen, a non-stimulated body part. This spread of ownership from the hand to the abdomen supports taking participants' reports of a sense of whole body ownership at face value.

If we accept that subjects in the full-body illusion experience a sense of ownership for a mannequin's whole body, we can ask how this relates to the sense of ownership of the body parts. De Vignemont (2022) frames the relevant question as follows: "Is it one and the same thing to experience a body part or the whole body as one's own? Or if we rephrase the question in empirical terms, do the rubber-hand illusion (RHI) and the full-body illusion (FBI) manipulate the same notion of ownership?" (p. 48). I will pursue a specific version of this question that explicitly concerns representations: Are the representations that ground the sense of ownership of the body parts the same as those that ground the sense of ownership of the whole body?

Two possibilities present themselves. First, they might be the same representations. According to this view, the sense of ownership of the whole body might result from a particular combination of the representations that ground the sense of ownership of the body parts. Thus, whatever representational explanation accounts for the rubber hand illusion, for example, could be extended to explain the full-body illusion.

The second possibility – which I will defend – is that the two senses of ownership are grounded in independent representations. No combination of representations that ground the senses of body part ownership can ground the sense of whole body ownership. In the following section, I present

Bermúdez (2017)'s account of bodily ownership and suggest that it successfully explains local ownership. In §4.3, I then argue that his account cannot adequately explain global ownership.

4.2 Bermúdez's account of bodily ownership

In this section, I present Bermúdez (2017)'s account of bodily ownership. I begin by explaining his proposal that we represent the space of the body using a hierarchy of generalized cones. I then show how this account explains body part ownership under normal conditions and in cases where ownership is disrupted, such as the rubber hand illusion and somatoparaphrenia.

Before presenting Bermúdez (2017)'s account, I should note an important difference in our target phenomena. While I have argued that empirical evidence points to a sense of bodily ownership as a genuine aspect of phenomenology, Bermúdez disagrees with this conclusion. According to him, "there are facts about the phenomenology of bodily awareness (about position sense, movement sense, and interoception) and there are judgments of ownership, but there is no additional feeling of ownership" (Bermúdez, 2011, p. 166). According to his view, there is no feeling of ownership over and above our awareness of bodily position, movement, and interoceptive sensations.

Given this, Bermúdez (2017) aims to account for judgments of ownership rather than feelings of ownership. However, this does not undermine the representational analysis I want to pursue. Judgments and purported feelings of ownership are thought to co-occur – indeed, judgments of ownership are typically taken as evidence for feelings of ownership. So, whether we understand ownership in terms of judgments or feelings, the same representational question arises: what are the underlying representations that make ownership (however construed) possible? I will take Bermúdez's theory to answer this broader representational question.

Bermúdez (2017) begins his account of bodily ownership by identifying two features that distinguish our experiences of bodily space from our experiences of extra-personal space: "Boundedness" and "Connectedness." To say that we experience the body as Bounded is to say that "bodily events are experienced within the experienced body (a circumscribed body-shaped volume whose boundaries define the limits of the self)" (p. 124). Bermúdez clarifies that while the boundaries of the experienced body are usually the boundaries of the physical body, the two can come apart. For example, sensations in a phantom limb are felt within the boundaries of the experienced body despite falling outside the physical body. Only the space of the body is experienced as Bounded; extra-personal space is not. To say that we experience the body as Connected is to say that "the spatial location of a bodily event is experienced relative to the disposition of the body as a whole" (p. 126). When I feel a pain in my right foot, for example, I do not merely experience the pain as in my right foot – I experience the pain relative to the space of the rest of my body. Because Connectedness presupposes Boundedness, it, too, distinguishes bodily space from extra-personal space.

For Bermúdez, Boundedness and Connectedness are features of bodily experience that a representational theory of bodily ownership should explain. He proposes that we represent the space of the body using a "hierarchy of generalized cones linked by mechanical joints" (p. 136). Each body part is represented as a generalized cone – a surface "generated by moving a cross-section along an axis, maintaining its shape but possibly varying its size" (p. 135). That one's body parts are represented as cones – i.e. bounded spaces – is intended to explain Boundedness. Connectedness is explained by the hierarchical organization of these cones. The hierarchy extends from the smallest body parts at the ends of the extremities (e.g., the last section of the finger following the outermost joint) all the way back to the torso. A bodily sensation is located relative to the limb in which it occurs using a cylindrical coordinate system whose origin falls on the joint immediately controlling movements of that limb; it is located relative to the body more generally by supplementing its limb-relative location "with a chain of relative joint angles that collectively specify the location of the given limb relative to the immovable torso" (p. 136). Each cone thus is linked to the others in the hierarchy, ensuring that sensations in any body part are experienced relative to the spatial organization of the whole body.

To see how this model of bodily space might account for body part ownership, I will work through specific cases. To begin, consider how the model can account for body part ownership under normal conditions. Under normal conditions, the model of the body should match the physical body. More specifically, the cones and spatial relations between them should match the body parts and spatial relations between them. In this way, we should expect that one has bodily ownership for their own physical body.

Now consider how the model might account for the rubber hand illusion, in which one experiences bodily ownership for a rubber hand rather than one's real hand. The model explains this shift as follows: the cone centered on one's wrist is misrepresented relative to the cone centered on one's elbow. The former matches the position of the rubber hand relative to the elbow rather than the position of one's real hand. As a result, one experiences bodily ownership for the rubber hand, rather than one's real hand.

The model can also account for limb disownership in somatoparaphrenia. Consider the somatoparaphrenic patient who experiences disownership of her paralyzed left hand and believes that it belongs to her niece. According to this account, her left wrist-centered cone is no longer represented; it is no longer part of her hierarchical model of her body. Because her hand representation has been excluded from her body representation, the patient no longer experiences bodily ownership for that hand.

Beyond explaining the rubber hand illusion and somatoparaphrenia, Bermúdez's model gains broader support from the patterns of ownership change in these cases. Bermúdez's hierarchy forms a tree structure because of how our body parts are structured. If bodily ownership is grounded in such a tree structure, we should expect specific propagation effects: changes to ownership of inner body parts (those closer to the torso) should always be accompanied by corresponding changes to outer body parts (those farther from the torso). For instance, if ownership of one's arm changes, ownership of the hand and fingers should change as well, since alterations to any branch of a tree structure affect all its offshoots.

The evidence appears to support this prediction. In the rubber hand illusion, ownership shifts from one's real hand to the rubber hand, and this shift extends to the fingers – subjects experience ownership not only of the rubber hand but also of its fingers. Similarly, in somatoparaphrenia, patients who lose ownership of their left hand also lose ownership of the hand's fingers. These patterns align with the tree structure model's predictions. In the rubber hand illusion, when the wrist-centered reference frame shifts position in the hierarchy, the finger-centered reference frames (as its offshoots) shift accordingly. In somatoparaphrenia, when the wrist-centered reference frame is dropped from the tree structure, the finger-centered reference frames are as well. This propagation from inner to outer body parts provides evidence that bodily ownership is grounded in hierarchical representations of the kind Bermúdez describes.

Thus far, we have seen how Bermúdez's hierarchical model successfully explains body part ownership. Bermúdez appears to think that this model can also explain whole body ownership. In his paper, he discusses research on full body ownership and expresses that his "proposal concerns the judgments of ownership that we make about our *bodies* and body parts" (p. 181, emphasis mine). That said, he does not work through how his account might apply to whole body ownership. Presumably, he takes whole body ownership to result from one's experience of the space of one's whole body as Bounded and Connected, just like body part ownership. One experiences the space of one's whole body as Bounded and Connected given how it is represented: the whole body is represented via a hierarchy of interconnected cones. That the cones represent a bounded volume makes one's experiences of whole body Bounded; that the cones are represented as connected to each other makes one's experiences of one's whole body Connected. In this way, the same representations that account for ownership of the body parts seem to account for ownership of the whole body.

This approach is theoretically parsimonious. It provides a single model of bodily representation that explains both body part ownership and whole body ownership. However, as I show in the next section, recent empirical evidence reveals a dissociation between the two kinds of ownership that challenges this approach.

4.3 A dissociation between global and local bodily ownership

In this section, I challenge the application of Bermúdez (2017)'s hierarchical model to global ownership by arguing for the following claim:

> *Representational Independence:* The representations that ground the sense of ownership for the body parts are not the same as the representations that ground the sense of ownership for the whole body.

We can understand this claim by imagining a Venn diagram with the representations grounding local ownership in one circle and those grounding global ownership in another. According to Representational Independence, these circles do not overlap – the representations that ground local ownership do not ground global ownership, and vice versa. Note that Representational Independence does not amount to or entail the probabilistic or causal independence of the two senses of bodily ownership. It is fully compatible with this view that local and global ownership experiences influence each other's likelihood, or that the use of one set of representations causally contributes to the use of the other. Such causal and probabilistic relations fall outside the claim's scope.

I now turn to review three sources of evidence that support Representational Independence: subjective reports, neuroimaging findings, and statistical analyses.

4.3.1 Evidence from subjective reports

O'Kane et al. (2024) found evidence of a bidirectional dissociation between the senses of local and global ownership. More specifically, they found evidence that individuals can experience an illusory sense of ownership for a mannequin's body part without experiencing a sense of ownership for the mannequin's whole body. At the same, they found that individuals can experience an illusory sense of ownership for a mannequin's whole body without experiencing a sense of ownership for all of its parts. This two-way dissociation between experiences of local and global ownership supports the idea that such experiences are grounded in distinct representations.

O'Kane et al. (2024)'s experimental setup mirrored that used by Petkova & Ehrsson (2008) but involved eight different experimental conditions. These varied with respect to whether synchronous or asynchronous strokes were applied to the participants' and mannequin's right arms, torsos, and right legs. For example, one experimental condition involved stroking all three body parts synchronously. Another involved synchronous strokes to the participant's and mannequin's right arms, and asynchronous strokes to their torsos and right legs. A third condition involved synchronous strokes to the participant's and mannequin's right arms and torsos, and asynchronous strokes to their right legs. Following the experiment, participants completed a questionnaire similar to that used by Botvinick & Cohen (1998) during the rubber-hand illusion and Petkova & Ehrsson (2008) during the full-body illusion. The questionnaire was comprised of ten statements, which participants were asked to score on a 7-point scale from "strongly disagree" to "strongly agree." One question assessed the sense of global ownership ("I felt as though the mannequin's whole body were my own body"), while others assessed the senses of local ownership ("I felt as though the mannequin's right arm were my arm").

The questionnaire results revealed a two-way dissociation between the senses of local and global ownership. When only one body part received synchronous stimulation and the other two body parts received asynchronous stimulation, participants reported a sense of ownership for the mannequin's synchronously stimulated body part but did not report a sense of ownership for the mannequin's whole body.[4] This suggests that subjects can experience local ownership without global ownership. By contrast, when two body parts received synchronous stimulation and the third received asynchronous stimulation, participants reported a sense of ownership for the mannequin's whole body, but did not report a sense of ownership for the mannequin's asynchronously stimulated

[4]This observation seems congruent with results from the rubber-hand illusion, where a rubber hand is experienced as owned in virtue of congruent multisensory stimulation to one's real hand and the rubber hand.

body part.[5] This suggests that subjects can experience global ownership without local ownership.

O'Kane et al. (2024)'s findings suggest that none of the representations that ground the sense of local ownership is individually necessary or sufficient to ground the sense of global ownership. As we have seen, one can experience a sense of global ownership for a mannequin's whole body without experiencing a sense of ownership for the mannequin's hand. This suggests that the representation that grounds the sense of ownership for one's hand is not necessary to ground an experience of global ownership. The same can be said for each of the representations that ground ownership of a body part. We have also seen that one can experience a sense of ownership for a mannequin's hand without experiencing a sense of ownership for its whole body. This suggests that the representation that grounds the sense of ownership for one's hand is not sufficient to ground the sense of global ownership.

Representational Independence, if true, would explain why none of the representations that ground the sense of local ownership is individually necessary or sufficient to ground the sense of global ownership. The representations that ground the sense of local ownership would be completely distinct from the representations that ground global ownership, and so, none of the former would be individually necessary or sufficient for the latter.

That said, Representational Independence is not the only interpretation of these findings. One might suggest that, depending on the context, various combinations of representations that ground the senses of local ownership are jointly sufficient to ground the sense of global ownership. In one context, the representations that ground the senses of ownership for the hand and torso may be sufficient to ground the sense of global ownership, while in another context, the representations that ground the senses of ownership for the torso and leg may be sufficient, and in a third context, the representations that ground the senses of ownership for the hand and leg. If something like this were right, then Representational Independence would be false. The representations grounding the sense of global ownership would be an amalgam of the representations grounding the senses of

[5] The researchers note that full-body ownership appeared "related to the number of synchronously stimulated parts, with the strongest illusion occurring in the fully synchronous condition" (O'Kane et al., 2024, p. 14).

local ownership, and so, the two sets of representations would not be independent.

Given the different possible interpretations of O'Kane et al. (2024)'s results, let us turn to consider another source of evidence that points in favor of Representational Independence.

4.3.2 Neural evidence

Neuroimaging evidence also supports Representational Independence. Petkova et al. (2011) used fMRI to analyze the activity of different brain regions during the full-body illusion. Three findings led them to conclude that "a genuine full-body ownership representation" – rather than a collection of body part representations – is instantiated in the left ventral premotor cortex (p. 1120). First, activity in this area was higher in experimental conditions in which the full-body illusion was elicited than in those in which it was not. In other words, such activity correlated with subjective full-body ownership ratings. Second, activity in the premotor cortex occurred irrespective of which body part received synchronous stimulation: activity occurred when the full-body illusion was elicited using synchronous stimulation to both the participant's and mannequin's i) hand and ii) abdomen. Such activity thus appeared insensitive to information about specific body parts. Third, this insensitivity to body part specific information was confirmed using multivoxel pattern analysis. Multivoxel pattern analysis allows for a more fine-grained understanding of neural activity patterns. Petkova et al. (2011) trained a classifier on activity patterns in the premotor cortex to detect when the full-body illusion was elicited given stimulation to the abdomen. They then found that this classifier could accurately detect when the illusion was elicited given stimulation to the hand. (They also found the reverse: a classifier trained to detect when the illusion was elicited given stimulation to the hand could accurately detect when the illusion was elicited given stimulation to the abdomen.) Thus, this fine-grained analysis confirmed that activity patterns in the premotor cortex did not carry information specific to a body part. Rather, as the researchers concluded, the results suggested "that activity in the left ventral premotor cortex reflects ownership generalized to the whole body" (p. 1120).[6]

The reviewed neuroimaging evidence supports the conclusion that global and local ownership are grounded in different neural activity patterns. If we assume that these different neural activity patterns instantiate different representations, then we have reason to think that global and local ownership are grounded in independent representations. In short, the neural evidence seems to support Representational Independence.

4.3.3 Statistical evidence

A third piece of evidence in favor of Representational Independence comes from a statistical analysis of the relation between local and global ownership from O'Kane & Ehrsson (2021). Following a modified version of the experiment performed in Petkova & Ehrsson (2008)'s study of the full-body illusion, O'Kane & Ehrsson (2021) had participants rate their agreement with statements concerning their feelings of ownership for a mannequin's 1) right arm, 2) left arm, 3) trunk, 4) right leg, 5) left leg, and 6) whole body. They then aggregated these ratings and performed a regression analysis of local ownership ratings on global ownership ratings. Their aim was to determine whether and how much local ownership ratings determine global ownership ratings. They found that "approximately half of the variance (57.5 percent) in illusory full-body ownership ratings could be attributed to the variance in illusory ownership for all body parts" (p. 17). This entails that full-body ownership ratings could not be predicted on the basis of body-part ownership ratings alone.

Representational Independence would explain why full-body ownership ratings cannot be predicted using body-part ownership ratings alone. If the representations that ground local and global ownership are independent, then no combination of the representations that ground local ownership is sufficient to generate experiences of global ownership. Therefore, local ownership ratings

[6] These results were replicated and expanded by Gentile et al. (2015), who found that activity patterns in the left ventral premotor cortex correlated with subjective global ownership ratings and were invariant to whether the stimulation which elicited the full-body ownership illusion was on the hand, abdomen, or leg.

cannot be used to predict global ownership ratings.

O'Kane & Ehrsson (2021) themselves articulate support of Representational Independence, albeit in slightly different terms. As they explain, their regression analysis results suggest that another variable (besides local ownership ratings) must be involved in the determination of global ownership ratings. They then write that "this variable likely reflects cognitive processes specific to the full-body ownership percept and involves neural substrates unique to those that represent the owned body parts" (p. 21). Cognitive processes *specific* to the full-body ownership percept would not be involved in local ownership percepts. O'Kane & Ehrsson (2021) suggest that such processes are instantiated in the neural substrates specific to global ownership (those previously examined). In this way, the researchers interpret their results as supporting the independence of global ownership from local ownership with respect to both cognitive processes and neural substrates.[7]

4.3.4 Implications for Bermúdez's account

Taken together, the three lines of evidence reviewed provide compelling support for Representational Independence. The behavioral evidence demonstrates that local and global ownership can dissociate in both directions. The neural evidence reveals that global ownership correlates with activity patterns in brain regions that are invariant to specific body part information. And finally, the statistical evidence shows that local ownership ratings cannot fully predict global ownership ratings. While each line of evidence has alternative interpretations, their convergence strengthens the case that the representations grounding local and global ownership are indeed independent.

If the empirical evidence supports Representational Independence, then Bermúdez's hierarchical model cannot adequately explain global ownership, despite its success with local ownership.

[7] Another possibility is that the missing variable(s) are additional body-part ownership ratings. O'Kane & Ehrsson (2021)'s questionnaire only covers some of the body parts (the two arms, the two legs, and the trunk), but misses others (the hands, feet, and head). Perhaps, adding these other ownership ratings would fully account for the unexplained variance in full-body ownership ratings. While this remains a possibility, it does not fit as well with the neuroimaging results, which suggest that local and global ownership are grounded in different neural substrates.

Representational Independence holds that local and global ownership require distinct representations, while Bermúdez's account grounds both in the same hierarchy of generalized cones. We thus seem to stand in need of a new representational account of global ownership.

4.4 A global ownership representation

What might such a new representational account look like? In this section, I will sketch one possibility. I propose that global ownership is grounded in a single representation of self. This representation operates according to rules that differ from those that govern local ownership. Most notably, use of the global ownership representation follows what I call the Constitution Rule, while use of the hierarchy of body part representations follows the Parthood Rule. I will argue that this representational account of global ownership differs from Bermúdez (2017)'s hierarchy of body part representations in three respects and is compatible with the empirical evidence for Representational Independence.

I propose that global ownership is grounded in a single representation that carries indexical relational content – specifically, the relation of being owned by the perceiver. This interpretation aligns with how global ownership is measured empirically: participants in full-body illusion studies agree to statements that use possessive language, such as "I felt as though the mannequin's whole body were my own body." This global ownership representation carries indexical information about the ownership relation without specifying any location or spatial properties of the perceiver. When combined with a representation of an object, it represents the object as owned by oneself, but considered alone, the global ownership representation lacks spatial content.

This representation is used in conjunction with complex multisensory representations of one's body and surrounding environment. Under normal conditions, it is applied to that object in the scene that corresponds to the perceiver's actual body, thereby marking that object as belonging to the perceiver. When global ownership is disrupted, as in the full-body illusion, the representation is incorrectly applied to an object that is not the perceiver's body.

When are objects represented as standing in the ownership relation to the perceiver? Given variations in the circumstances under which individuals have illusory experiences of global ownership, we can posit certain rules governing use of the global ownership representation. For example, an illusory experience of owning a mannequin's body seems to require that one feels touches on one's own body at the same time as one sees touches applied to the mannequin's body. When these cues are asynchronous, the full-body illusion disappears (Petkova & Ehrsson, 2008). Temporal synchrony of multisensory cues thus seem to constrain the conditions under which the global ownership representation is applied. Similarly, the strength of the full-body illusion appears to vary depending on whether the visually observed object looks like a human (Petkova & Ehrsson, 2008). This suggests that the illusion might be constrained by a humanoid shape rule, according to which the closer an object's shape is to that of the human body, the more likely (all else equal) that the object is represented as belonging to one's self. Other possible rules likely concern the object's position relative to oneself (Maselli & Slater, 2014) and congruence between visually observed movements of the object and proprioceptively felt movements of one's body (Banakou et al., 2013).

Empirical evidence suggests that many rules governing global ownership also govern local ownership. For example, the rubber hand illusion, like the full-body illusion, appears to depend on synchronous multisensory cues delivered to the rubber hand and one's own hand (Botvinick & Cohen, 1998). Likewise, the illusion depends on whether the perceived object's shape resembles the shape of one's own hand (Haans et al., 2008). This suggests that the constraints that affect whether objects are incorporated into the hierarchy of body part representations are similar to those that affect whether something is represented as one's whole body.

Despite these similarities, the rules governing local and global ownership appear to differ in one crucial respect. Local ownership appears governed by what I call the Parthood Rule: an object is more likely to be represented as part of the hierarchy of body part representations if (all else equal) the whole of which the object is a part is represented as belonging to one's self.

For example, one is more likely to represent a mannequin's hand as part of the hierarchy if one represents the mannequin's whole body as belonging to one's self. This explains the spread of feelings of ownership from whole bodies to body parts found in Petkova & Ehrsson (2008)'s study.

By contrast, global ownership appears governed by what I will call the Constitution Rule: the greater the number of objects represented in the hierarchy of body part representations, the more likely (all else equal) that the whole that the objects compose is represented as belonging to one's self. This rule explains O'Kane et al. (2024)'s findings. Synchronous stimulation of only one of the mannequin's body parts produced local ownership for the part but was insufficient, given the Constitution Rule, to trigger global ownership for the mannequin's whole body. However, synchronous stimulation of two body parts was sufficient to trigger global ownership for the mannequin's whole body.[8]

We can now consider the three ways in which the global ownership representation differs from Bermúdez's hierarchy of body part representations. First, it differs in number: many representations ground the many senses of local ownership – one for each of the different body parts – while only a single representation grounds the sense of global ownership. Second, it differs in content: the hierarchy of generalized cones that Bermúdez posits has inherently spatial content, with each cone representing a certain bounded volume, while the global ownership representation carries only indexical relational content without any spatial content. Third, as we have seen, it differs in one of the rules governing its application: local ownership is constrained by the Parthood Rule, while global ownership is constrained by the Constitution Rule.

This representational account of global ownership is not only theoretically distinct from Bermúdez's approach – it also aligns with the empirical evidence for Representational Independence. Specifically, it is consistent with behavioral evidence showing that subjective reports of local and global ownership can come apart, with neural evidence that global ownership is instantiated in different brain regions than local ownership, and with statistical evidence that local ownership ratings cannot

[8]This discussion of the rules governing the use of local and global ownership representations is inspired by O'Kane et al. (2024)'s Bayesian hierarchical model of local and global ownership.

fully predict global ownership ratings.

4.5 Towards the conceptual self representation

I want to close by suggesting that the global ownership representation occupies a middle ground between the body part representations used in perception and the conceptual representation of self used in *de se* thought. This follows from the preceding comparison of the global ownership representation with the body part representations and the following comparison of the global ownership representation with the conceptual self-representation.

Two of the features that distinguish the global ownership representation from the hierarchy of body part representations mark points of similarity between it and the conceptual representation of self. These are number and content. With respect to number, both are single representations rather than hierarchies of multiple representations. With respect to content, both lack spatial information when considered in isolation.

Of course, the global ownership representation is not identical to the conceptual self representation. The key difference lies in what they represent. While both carry indexical information about the representing individual, the global ownership representation carries relational content ('owned by me'), whereas the conceptual self-representation carries individual content ('me'). The former is used to represent objects as standing in a relation to the self; the latter is used to represent the self directly.

These comparisons suggest that the global ownership representation occupies an intermediate position between the many representations of the body parts used in perception and the self-representation found in thought. Determining the full path from this relational self-representation to the pure conceptual self-representation remains for future work. What should be clear is that understanding this transition will likely require tracing how different kinds of representations of self interact throughout the mind.

Bibliography

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.

Alsmith, A. J. (2021). Bodily structure and body representation. *Synthese*, *198*(3), 2193–2222.

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.

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

Andersen, R. A., & Zipser, D. (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.

Armel, K. C., & Ramachandran, V. S. (2003). Projecting sensations to external objects: Evidence from skin conductance response. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1523), 1499–1506.

Ávila, I. (2014). Evans on bodily awareness and perceptual self-location. *European Journal of Philosophy*, *22*(2), 269–287.

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

Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, *110*(31), 12846–12851.

Barendregt, M., Harvey, B. M., Rokers, B., & Dumoulin, S. O. (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. (2011). Bodily awareness and self-consciousness. In S. Gallagher (Ed.) *Oxford handbook of the self*, (pp. 157–179). Oxford University Press.

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.

Billon, A. (2017). Mineness first: Three challenges to the recent theories of the sense of bodily ownership. In F. de Vignemont, & A. J. Alsmith (Eds.) *The subject's matter: Self-consciousness and the body*, (pp. 189–216). MIT Press.

Blanke, O., Landis, T., Spinelli, L., & Seeck, M. (2004). Out-of-body experience and autoscopy of neurological origin. *Brain*, *127*(2), 243–258.

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends in Cognitive Sciences*, *13*(1), 7–13.

Blanke, O., & Mohr, C. (2005). Out-of-body experience, heautoscopy, and autoscopic hallucination of neurological origin: Implications for neurocognitive mechanisms of corporeal awareness and self-consciousness. *Brain Research Reviews*, *50*(1), 184–199.

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

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756–756.

Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge: Cambridge, MA: Harvard 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.

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

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

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

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

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

Burge, T. (2014). Perception: Where mind begins. *Philosophy*, *89*(3), 385–403.

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.

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

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

Cassam, Q. (1997). *Self and world*. Clarendon Press.

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.

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

Colby, C. L., & Duhamel, J.-R. (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.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, *60*(23), 685.

de Vignemont, F. (2018). *Mind the body: An exploration of bodily self-awareness*. Oxford University Press.

de Vignemont, F. (2022). Ten problems of bodily ownership. In A. J. Alsmith, & M. R. Longo (Eds.) *The Routledge handbook of bodily awareness*, (pp. 46–57). Routledge.

Ehrsson, H. H., Wiech, K., Weiskopf, N., Dolan, R. J., & Passingham, R. E. (2007). Threatening a rubber hand that you feel is yours elicits a cortical anxiety response. *Proceedings of the National Academy of Sciences*, *104*(23), 9828–9833.

Ekstrom, A. D., Huffman, D. J., & Starrett, M. (2017). Interacting networks of brain regions underlie human spatial navigation: A review and novel synthesis of the literature. *Journal of Neurophysiology*, *118*(6), 3328–3344.

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

Erkelens, C. J., & van Ee, R. (2002). The role of the cyclopean eye in vision: Sometimes inappropriate, always irrelevant. *Vision Research*, *42*(9), 1157–1163.

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

Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*. MIT press.

Galletti, C., Battaglini, P. P., & Fattori, P. (1993). Parietal neurons encoding spatial locations in craniotopic coordinates. *Experimental Brain Research*, *96*, 221–229.

Gallistel, C. (1999). Coordinate transformations in the genesis of directed action. In B. O. M. Bly, & D. E. Rummelhart (Eds.) *Cognitive science*, (pp. 1–42). Elsevier.

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.

Gandola, M., Invernizzi, P., Sedda, A., Ferrè, E. R., Sterzi, R., Sberna, M., Paulesu, E., & Bottini, G. (2012). An anatomical account of somatoparaphrenia. *Cortex*, *48*(9), 1165–1178.

Gentile, G., Björnsdotter, M., Petkova, V. I., Abdulkarim, Z., & Ehrsson, H. H. (2015). Patterns of neural activity in the human ventral premotor cortex reflect a whole-body multisensory percept. *Neuroimage*, *109*, 328–340.

Gerstmann, J. (1942). Problem of imperception of disease and of impaired body territories with organic lesions: Relation to body scheme and its disorders. *Archives of Neurology & Psychiatry*, *48*(6), 890–913.

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

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

Grossberg, S., Guenther, F., Bullock, D., & Greve, D. (1993). Neural representations for sensory-motor control, II: Learning a head-centered visuomotor representation of 3-D target position. *Neural Networks*, *6*(1), 43–67.

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

Haans, A., Ijsselsteijn, W. A., & de Kort, Y. A. W. (2008). The effect of similarities in skin texture and hand shape on perceived ownership of a fake limb. *Body Image*, *5*(4), 389–394.

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").

Holmes, N. P., Snijders, H. J., & Spence, C. (2006). Reaching with alien limbs: Visual exposure to prosthetic hands in a mirror biases proprioception without accompanying illusions of ownership. *Perception & Psychophysics*, *68*(4), 685–701.

Hume, D. (1978). *A treatise of human nature*. (L. A. Selby-Bigge and P. H. Nidditch, Eds.) Clarendon Press. Original work published in 1739.

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.

Johnson, G. M. (2020). The structure of bias. *Mind*, *129*(516), 1193–1236.

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

Karnath, H., Schenkel, P., & Fischer, B. (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.

Kaufmann, A. (2015). Animal mental action: Planning among chimpanzees. *Review of Philosophy and Psychology*, *6*, 745–760.

Keil, F. (2013). *Developmental psychology: The growth of mind and behavior*. WW Norton & Company.

Klatzky, R. L. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel, & K. F. Wender (Eds.) *Spatial cognition: An interdisciplinary approach to representing and processing spatial knowledge*, (pp. 1–17). Springer Berlin Heidelberg.

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

Li, D., Karnath, H.-O., & Rorden, C. (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).

Maselli, A., & Slater, M. (2014). Sliding perspectives: Dissociating ownership from self-location during full body illusions in virtual reality. *Frontiers in Human Neuroscience*, *8*, 693.

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.

Moseley, G. L., Olthof, N., Venema, A., Don, S., Wijers, M., Gallace, A., & Spence, C. (2008). Psychologically induced cooling of a specific body part caused by the illusory ownership of an artificial counterpart. *Proceedings of the National Academy of Sciences*, *105*(35), 13169–13173.

Mullette-Gillman, O. A., Cohen, Y. E., & Groh, J. M. (2005). Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. *Journal of Neurophysiology*, *94*(4), 2331–2352.

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

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

O'Kane, S. H., Chancel, M., & Ehrsson, H. H. (2024). Hierarchical and dynamic relationships between body part ownership and full-body ownership. *Cognition*, *246*, 105697.

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

Ono, H., Mapp, A. P., & Howard, I. P. (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.

Orbán, K., & Wong, H. Y. (2023). The sense of body ownership. In M. García-Carpintero, & M. Guillot (Eds.) *Self-experience: Essays on inner awareness*, (pp. 142–165). Oxford University Press.

O'Kane, S. H., & Ehrsson, H. H. (2021). The contribution of stimulating multiple body parts simultaneously to the illusion of owning an entire artificial body. *PLOS One*, *16*(1), e0233243.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT press.

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*. MIT Press.

Peacocke, C. (1999). *Being known*. Clarendon Press.

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

Petkova, V. I., Björnsdotter, M., Gentile, G., Jonsson, T., Li, T.-Q., & Ehrsson, H. H. (2011). From part- to whole-body ownership in the multisensory brain. *Current Biology*, *21*(13), 1118–1122.

Petkova, V. I., & Ehrsson, H. H. (2008). If I were you: Perceptual illusion of body swapping. *PLOS One*, *3*(12), e3832.

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*. 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.

Rohde, M., Wold, A., Karnath, H.-O., & Ernst, M. O. (2013). The human touch: Skin temperature during the rubber hand illusion in manual and automated stroking procedures. *PLOS One*, *8*(11), e80688.

Saleem, A. B., Diamanti, E. M., Fournier, J., Harris, K. D., & Carandini, M. (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.

Sebo, J. (2017). Agency and moral status. *Journal of Moral Philosophy*, *14*(1), 1–22.

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.

Siciliano, B., Khatib, O., & Kröger, T. (2008). *Springer handbook of robotics*, vol. 200. Springer.

Sun, L. D., & Goldberg, M. E. (2016). Corollary discharge and oculomotor proprioception: Cortical mechanisms for spatially accurate vision. *Annual review of vision science*, *2*(1), 61–84.

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

Velleman, J. D. (1992). What happens when someone acts? *Mind*, *101*(403), 461–481.

Wandell, B. A., & Winawer, J. (2011). Imaging retinotopic maps in the human brain. *Vision Research*, *51*(7), 718–737.

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. (1922). *Tractatus logico-philosophicus*. (C.K. Ogden, Trans.). Routledge & Kegan Paul.

Wittgenstein, L. (1958). *The blue and brown books*. Oxford: Blackwell.

Zaehle, T., Jordan, K., Wüstenberg, T., Baudewig, J., Dechent, P., & Mast, F. W. (2007). The neural basis of the egocentric and allocentric spatial frame of reference. *Brain Research*, *1137*, 92–103.

Zahavi, D., & Kriegel, U. (2015). For-me-ness: What it is and what it is not. In D. O. Dahlstrom, A. Elpidorou, & W. Hopp (Eds.) *Philosophy of mind and phenomenology*, (pp. 36–54). Routledge.

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.