

Locating the Self: Indexicals in Perception and Action

Catherine Hochman

Abstract

Our actions are guided by, and thus partially explained by, our perceptions. Accordingly, theories of perceptual content must be compatible with action explanations. It is commonly assumed that the dominant view, which holds that perceptual representations include *de se* (first-personal) contents, readily supports such explanations. More recently, a simpler alternative has been proposed: that perceptual representations carry merely indexical locational contents that refer to a location, that occupied by the perceiver. Proponents of this view contend that it, too, can be used in explanations of action. In this paper, I argue that both views capture something essential that the other neglects. *De se* contents are well-suited to explain the motivational aspect of action, while indexical locational contents can explain the coordination between perception and action. Yet taken individually, each view faces explanatory challenges. To address these, I propose the Located Individual View, which holds that perceptual representations include both *de se* and indexical locational contents. I contend that this hybrid account can best explain how perception both motivates and guides action.

Keywords: perceptual content, *de se* representation, indexical content, action explanation

1 Introduction

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 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).

¹See 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.

²That 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 §4 and §5.

³Whether 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.

In this paper, 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 §2, 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 §3, 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 §4, I discuss how the view that posits *de se* contents struggles to satisfy the first desideratum. Finally, in §5, I articulate the Located Individual View and argue that it satisfies both desiderata.

2 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

⁴A related debate 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.

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 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

⁵Kaplan (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).

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.

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.

3 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 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 that provides the motivational force

⁶More 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).

⁷For 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 merely causal.

One might wonder whether my arguments extend to non-human animals, given skepticism about whether belief-desire models

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 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.

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.

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 $\langle x \text{ moving towards } h \rangle$, one automatically forms an intention with the content $\langle \text{move left of } h \text{ to avoid } x \rangle$. 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 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 §2). 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.

⁸It 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.

4 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 §5.

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 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

⁹Answers 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.

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 (§2). 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 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

¹⁰For 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).

¹¹Skepticism 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).

right, 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).

5 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 $\langle i, h \rangle$. 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 “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 (§4).

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 §4).

Now, there is no need to posit use of the bridge premise ‘*I am here*’ to connect one’s perceptual belief

¹²One 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).

¹³By 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.

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.

6 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 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.

References

- Alsmith, A. (2017). Perspectival structure and agentive self-location. In F. de Vignemont, & A. J. T. Alsmith (Eds.) *The Subject's Matter: Self-Consciousness and the Body*, (pp. 263–288). MIT Press.
- Alsmith, A. J. (2021). Bodily structure and body representation. *Synthese*, 198(3), 2193–2222.
- Ávila, I. (2014). Evans on bodily awareness and perceptual self-location. *European Journal of Philosophy*, 22(2), 269–287.
- Bermúdez, J. L. (1998). *The paradox of self-consciousness*. MIT Press.
- Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge: Cambridge, MA: Harvard University Press.
- 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.
- Burge, T. (2010). *Origins of objectivity*. Oxford University Press.
- Campbell, J. (1994). *Past, space, and self*. MIT Press.
- Cassam, Q. (1997). *Self and world*. Clarendon Press.
- Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, 60(23), 685.
- Evans, G. (1982). *The varieties of reference*. (J. McDowell, Ed.). Clarendon Press; Oxford University Press.

- Fernandez Velasco, P. (2024). Self-location in perceptual experience: A top-down account. *Mind & Language*.
- 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 *Cognitive science*, (pp. 1–42). Elsevier.
- Groh, J. M. (2014). *Making space: How the brain knows where things are*. Harvard University Press.
- Hurley, S. L. (1998). *Consciousness in action*. Harvard University Press.
- Julesz, B. (1971). *Foundations of cyclopean perception*. University of Chicago Press.
- Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.) *Themes from Kaplan*, (pp. 481–563). Oxford University Press.
- 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.
- Mitchell, J. (2021). Self-locating content in visual experience and the ‘here-replacement’ account. *The Journal of Philosophy*, *118*(4), 188–213.
- 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.
- 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.
- 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*, (pp. 3–21).
- Perry, J. (1986). Thought without representation. *Proceedings of the Aristotelian Society, Supplementary volumes*, *60*, 137–151.
- 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.

- 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.
- 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.
- 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.