



**UNIVERSITY OF LEEDS**

This is a repository copy of *The Inside-Out Binding Problem*.

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/144777/>

Version: Accepted Version

---

**Book Section:**

Salje, L (2019) The Inside-Out Binding Problem. In: Cheng, T, Deroy, O and Spence, C, (eds.) *Spatial Senses: Philosophy of Perception in an Age of Science*. Routledge Studies in Contemporary Philosophy . Routledge , Abingdon, Oxon, UK , pp. 307-329. ISBN 9781138506411

<https://doi.org/10.4324/9781315146935>

---

© 2019 Taylor & Francis. This is an author produced version of a chapter published in *Spatial Senses: Philosophy of Perception in an Age of Science*. Uploaded in accordance with the publisher's self-archiving policy.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

# 16

## The Inside-Out Binding Problem

*Léa Salje*

### 0. Introduction

In “The Refutation of Idealism” Kant argues that awareness of oneself as an object spatially located among others is a necessary condition on a certain kind of self-consciousness (B274–279). The argument, in (very) short, runs as follows. Through what Kant calls *empirical self-consciousness* I am aware of my mental states as ordered in time. Any temporal ordering, however, must be measured against an external reference point. So, the argument goes, empirical self-consciousness requires awareness of objects persisting outside me, which in turn requires awareness of myself as something that can bear spatial relations to other objects. So while empirical self-consciousness is direct consciousness of one’s existence as determined in time, it requires awareness of oneself as located in space.

The idea that there is something about self-conscious thought that requires a capacity to think of ourselves, *as* ourselves, as objects spatially located among others is a theme that finds echoes across the work of P. F. Strawson, Thomas Nagel, Gareth Evans, and more recently John Campbell, Quassim Cassam, and J. L. Bermúdez—variously developed, but always returning to the basic idea that there are demands on self-conscious thought that require us to conceive of ourselves in a *de se* way as objects spatially located among others. How might we develop such a self-conception? One plausible suggestion is by *perceiving* oneself as such. That is, by bodily self-perception.

For the animals that we are, the modes of bodily self-perception can be grouped in two: the *exteroceptive* senses, like vision, touch, and smell, through which one’s body is perceived as one object among many in the world; and the *interoceptive* senses, like proprioception (positional-sense), kinaesthesia (movement-sense), and nociception (pain-sense), through which a subject can perceive her own body and its parts only.

An intuition-friendly way of thinking of this is as a division between ways of perceiving our bodies “from the inside” and “from the outside.” This heuristic is fine as far as it goes, but clearly shouldn’t be taken overly

seriously. Inside/outside what? Humans are not bounded by skin all over—our outer edges are punctuated by nails, hair, eyeballs, and various orifices. Even heavily edited to deal with these glitches, a purely spatial criterion seems unpromising: given that our bodies are by and large topological donut-shaped figures, any way of drawing the boundaries between inside and outside is bound to seem a little arbitrary. (How, for example, should one draw a principled line in the case of the mouth, tongue, and throat?) More importantly, the inside/outside distinction does not really capture what we're after. With a bit of cooperation from you, I can easily look inside your eyelid, and with the right machinery there are few limits on visual exploration of our internal organs. But surely vision is one of the “outside” senses if anything is. In place of a spatial distinction, I characterize the divide in epistemic terms. On one side are the *single-object dedicated* (or body-exclusive) perceptual modalities through which one can *directly* gain knowledge about one's own body only; on the other are the *multiple-object* (or non-body-exclusive) modalities that can in principle provide perceptual information about more than one object. I will call the former the *interoceptive* senses, the latter *exteroceptive*.

This way of categorizing the senses tracks a stable difference in the ways of thinking about oneself directly sustained by their exercise. Through the interoceptive senses my body is directly sensorily presented to me *as me*, under a *de se* guise or mode of presentation. This “directness” consists in the fact that in order to form a first-person interoception-based judgment, like *I am hungry* or *my arms are crossed*, I need not draw on any further identification between the body (or body part) so perceived and myself, thought of in a *de se* way—the deliverances of the interoceptive senses, rather, are apt for unmediated uptake into a first-person thought. A symptom of this is that first-person interoceptive judgments are immune to error through misidentification (IEM) relative to the use of the first-person concept, which is to say that there is no possibility that a judgment, so formed, could be in error solely by virtue of a misidentification of the object known on that basis to be instantiating the perceived property with oneself, thought of first personally. There is no possibility of such an error in these cases, because the object so perceived is already directly perceived as oneself, so there is no question of having to *identify* it as such. Regardless of whether Shoemaker was right that “the identification of a presented object as oneself would have to go together with the possibility of misidentification,” the converse is certainly true: where there is no identification there can be no possibility of a misidentification (1968, p. 561).

The exteroceptive senses are not like this. I can know perfectly well by smell that someone smells a certain way, but misidentify the source of the smell as myself—or looking down at our intertwined fingers, I could know on that basis that *someone* has a hairy knuckle on one of their middle fingers, only mistakenly judge it to be me. All of this is just to say that first-person

judgments concerning one's own body made on the basis of smell, touch, vision, and hearing are not IEM relative to uses of the first-person concept. That's because unlike the interoceptive senses, they don't directly present their objects in a *de se* way, but rather as objects in the world among others, that must be identified as oneself in the course of forming first-person judgments on their basis.

This point of contrast between the two groups of senses creates a special difficulty for the foregoing suggestion that we develop our *de se* conceptions of ourselves as objects spatially located among others by straightforwardly perceiving ourselves as such. We have just seen that through the interoceptive senses we are perceptually acquainted with an object under a *de se* mode of presentation, but not as one object among others. And through the exteroceptive senses we are perceptually acquainted with a particular bodily object as one object among others in the world, but not directly *as oneself*. Now, a *de se* conception of oneself as an object spatially located among others encompasses both of these ways of thinking of oneself at once—that is, it depends on reconciling these two modes of presentation as determining the same referent, *me, this object spatially located among others*.

The difficulty introduced by contrasting these two sets of senses, then, is this. Suppose that it's right that we develop these self-conceptions by endorsing the contents of bodily self-perception. And suppose too that it's right that these two modes of self-perception respectively sustain self-directed thought under only one of these modes of presentation or the other. Then it is surely the case that the inputs from interoception and exteroception are integrated in episodes of bodily self-perception such that we perceive ourselves as the targets of both forms of perception at once. If that's right, then it's no great mystery that we could form *de se* conceptions of ourselves as objects located among others by endorsing the contents of *de se* perceptions of ourselves as objects located among others; we are able to *think* of ourselves this way because we *perceive* ourselves this way. Broadly stated, this chapter asks how this integration happens.

In the next section (§1) I make this question more precise by formulating it as a question about overlap in perceived properties across the two sets of senses. An initially tempting answer is that there is an overlap in perceived *spatial* properties at a time that, at least in part, grounds our awareness of the sameness of the body perceived through interoception and exteroception. I formulate this suggestion in §2, and raise a case against it in §3. In its place, I sketch the beginnings of an alternative answer that I take to avoid these problems—that is, that it is perceived *temporal* convergence that, at least in part, underpins our capacity to integrate inputs through the two sets of senses (§4). §5 raises some empirical predictions from these philosophical findings, and §6 concludes.

## 1. Refining the Question

## 16 The Inside-Out Binding Problem

To recap: through interoception one perceives a body as one's own, while through exteroception one perceives a body as a worldly object among others. If our *de se* self-conceptions as objects spatially located among others are developed by endorsing the contents of bodily self-perception, then in bodily self-perception one must be sensitive to the identity of the body (or body parts) interoceptively perceived *as one's own*, and exteroceptively perceived *as an object spatially located among others*. The question is, what underpins this perceptual sensitivity?

This isn't supposed to be a sceptical challenge. We clearly do normally succeed in so integrating interoceptive and exteroceptive contents, and we normally do so with minimal effort or attention—if we didn't, it would be a miracle that we could coordinate ourselves to get out of bed in the morning. But there is at least a *prima facie* puzzle about how it is done. The source of this puzzle is that the range of properties known about through the interoceptive senses seems very different from those known through exteroception. My sunburnt arm, for instance, is presented to me through vision via an array of shape and colour properties (swollen, reddish), while through nociception I am presented with sensation or pain properties. These differences give *prima facie* teeth to the question of what grounds my sensitivity to the sameness of the body part in which I know these diverse and non-overlapping properties to inhere through the two sets of senses.

In Quassim Cassam's (2007) helpful terminology, rather than a sceptical question, what we have here is a "how-possible" question that gets its force from the presence of an apparent obstacle. How is it possible that we are able to perceptually recognize the identity of a body (and its parts) perceived through interoception and exteroception, given the apparent absence of property overlap associated with the two sets of sense modalities? Cassam's framework offers us two strategies here. Either *overcome* the obstacle by showing that we don't need property overlap to have perceptual awareness of the identity of the bodily object given through the two sets of senses; or else *dissolve* it, by showing that there *is*, in fact, an overlap of the right kind to draw on. The first option, in this case, looks discouraging. How could one be aware of the identity of an object crossmodally perceived, absent any common properties represented across the different modalities? On what could one pin the identity? Of course, that it seems unpromising doesn't mean that it can't be done—but it is reason enough to prefer the second option as our starting point. This chapter pursues the second option.

Our task will push us down to the subpersonal level. After all, reflection on one's own present case will presumably be argument enough that in a typical personal level interoceptive-exteroceptive crossmodal perceptual state, one's body and its parts already present themselves as unified targets of crossmodal perception. Right now, for instance, I seem to perceive my typing hands as unified objects perceived both "from the inside" and "from the outside." But if this is right, then our how-possible question cannot have proper application

## 16 The Inside-Out Binding Problem

at the personal level—the question how I manage to consciously integrate individual modality-specific sensory inputs can't be well posed if that isn't something I routinely do (or that routinely happens to me) at a personal level. The question, rather, is how our subpersonal perceptual processing systems are able to affect this integration in a way that results in unified crossmodal personal-level perceptual experiences.

This question parallels the “binding” question in vision science. How do our visual systems succeed in binding colour, shape, surface reflectance, and other such cues in the visual field into personal-level conscious visual experiences as of unified objects? There are also crossmodal binding questions in the exteroceptive field about how perceptual cues coming from different exteroceptive sense modalities are appropriately bound together. How, for instance, do my relevant systems process incoming content from both vision and touch to produce my current unified experience of the keyboard in front of me as a single object that is both seen and felt?

To solve the exteroceptive crossmodal binding problem, my subpersonal mechanisms must find a way to overcome differences in the kinds of properties accessed via the two sense modalities. (I see the keyboard as black and rectangular; I feel its surface as cold and smooth. What binds together *this* patch of blackness with *this* patch of smoothness as coming from a single source?) In neo-Fregean terms this is a question posed at the level of sense about how to reconcile perceptual contents such that the sameness of their referent is recognized.<sup>9</sup> In terms more familiar to cognitive science, Anne Treisman explains,

Sensory information arrives in parallel as a variety of heterogeneous hints, (shapes, colours, motions, smells, and sounds) encoded in partly modular systems. Typically many objects are present at once. The result is an urgent case of what has been labelled the binding problem. We must collect the hints [and] bind them into the right spatial and temporal bundles.

(2003, p. 97)

Our present question likewise asks how, given the heterogeneity of “hints” coming in through the different modalities, our subpersonal mechanisms manage to bind inputs from interoceptive and exteroceptive sources to produce personal-level perceptual experiences as of single body parts perceived in both ways at once. It will be answered by finding a site, or sites, of property overlap between the two sets of senses to feature in a binding heuristic—that is, a principle used to guide the resolution of these interoceptive-exteroceptive binding tasks.

To dissolve our way out of the how-possible question earlier, then, our task is to find out what sort of binding principle, mentioning what sort of property overlap, is made use of by the relevant subpersonal mechanisms to resolve the

interoceptive-exteroceptive binding problem. This is an empirical question. But I think that there are distinctively philosophical considerations that can contribute to its answer.

## 2. Represented Spatial Overlap at a Time

Contrasting our interoceptive-exteroceptive case with the intramodal and the exteroceptive crossmodal cases just mentioned is an instructive place to begin our search. A standard answer in the case of these more familiar binding questions, in both the empirical and the philosophical literatures, appeals to a principle of spatial coincidence. In his contribution to this volume, for instance, Charles Spence writes,

Many authors have claimed that spatial co-location facilitates multisensory integration in humans. What is more, those making such claims typically do so without any caveats, or even any hint that the assertion might not, or at least not always, be true. It is safe to say that the spatial rule is, by now, firmly engrained in the cognitive neuroscience literature.

(p. X)<sup>10</sup>

Or to give an example from the philosophical literature, John Campbell writes of the visual intramodal case:

If the visual processing streams contain at least implicit information about the locations of the features being found—just where in the environment a particular colour or shape is, for example—then there is one effective strategy available. *Features found at the same location could be put together as features of a single object.*

(2002, p. 31, emphasis added)

Is a similar principle at work in the case of bodily self-perception? The suggestion, I take it, is that just as it is plausibly experienced spatial overlap between the colour and shape cues coming in through my visual system that dominantly ordinarily underwrites my conscious visual experience as of a unified black rectangular keyboard in front of me, there is experienced spatial overlap between the interoceptive and exteroceptive cues coming from my typing hands that typically explains my personal-level conscious crossmodal experience as of two unified body parts being perceived in both ways at once.

This is an initially credible suggestion on a couple of fronts. For one thing, our bodies are concrete worldly objects like any other, so there's no reason to think that what goes for non-bodily objects doesn't go for bodily objects too. In Evans's terms, the fundamental ground of difference that individuates our bodies is the very same fundamental ground of difference that individuates

## 16 The Inside-Out Binding Problem

any other concrete object: namely, its position in space and time (1982, p. 221). So if we have crossmodal representation of spatial overlap at a time, this surely gives us just as powerful an explanation of our crossmodal conscious experiences of unified body parts as in the case of any other crossmodally perceived worldly object.

The second front is that spatial content really is something that we get in abundance in both interoception and exteroception. The most spatially saturated exteroceptive senses are vision and touch, through which we get richly detailed and highly determinate information about the spatial profile of objects attentionally seen or felt (direction, distance, inter-object relative positions, arrangement of parts, etc.). Though less determinate, we typically also gain directional and distance information in audition. Likewise in interoception, spatial cues abound. If I feel a pain through nociception, I can easily tell on that basis whether it is located nearer the crown of my head or the soles of my feet; through proprioception, I can use cues from joint angles and muscle stretch to tell what relative positions my limbs are in; through kinaesthesia I can tell whether I am swinging my arm forwards or backwards relative to my trunk; and so on.

Given that (a) we plainly get plenty of spatial information through both sets of senses, and (b) there is good reason to think that represented spatial overlap at a time is just as good a guide to the sameness of object in the case of bodies as in the case of other worldly objects, a heuristic involving mention of spatial overlap at a time seems like a promising place to start:

*Spatial hypothesis.* Personal-level recognition of identity of body(-part) perceived both interoceptively and exteroceptively is made possible by subpersonal binding of cues from both forms of perception grounded on recognised co-location at a time of body(-part) perceived both ways at once.

Does the spatial hypothesis give us our answer to the how-possible question from §1?

### 3. The Problem With Represented Spatial Overlap at a Time

To see the difficulties with the spatial hypothesis as an answer to our question it will be helpful to consider Evans's response to Molyneux's question. That question, asked by William Molyneux in a letter to John Locke in 1688, asked whether a born-blind subject able to competently distinguish particular shapes through touch would have the capacity to visually recognize those shapes by vision alone upon the recovery of sight. Evans argued that the subject would, on the basis that the geometric properties of shapes (a 2D square, say) are given in a single egocentric frame of reference through both vision and touch. If the same spatial information about a shape is given through the two senses



## 16 The Inside-Out Binding Problem

in the same frame of reference, the idea is, then, that our shape concepts must evidently be crossmodally univocal.

Much of Evans's case for this affirmative answer to Molyneux's question is devoted to establishing the existence of a single frame of reference shared by vision and touch (and audition). He does this by appeal to the behavioural affordances couched in perceptual states: there is only one frame of reference because, as Evans puts it, "there is only one behavioural space" (1985, p. 390). If I am to make a grab for a perceptually presented rattle, I must move my hand in the same way and to the same position regardless of whether the rattle is seen, felt, or merely heard in the dark.<sup>13</sup>

This way of establishing the singularity of the crossmodal frame of reference also serves to establish that this single frame must be egocentrically structured. That's because the singularity is determined by the *shared significance of representations of space for action* across the modalities, and the way of representing space as used in action requires specifications of locations *relative to the actor's body*. After all, it won't help me act on the rattle to perceptually represent it as being located somewhere-or-other, or in an allocentric reference frame centred on some distinct, non-bodily object in the local environment. Rather, if these spatial contents are to guide my attempt to grab the rattle, then I must represent it as located in a position *relative to me and my grasping hand*. The picture we are left with is a crossmodally shared frame of reference, anchored to a single behavioural space, canonically specified in egocentric terms:

We envisage specifications like this: he hears the sound *up*, or *down*, *to the right* or *to the left*, *in front* or *behind*, or *over there*. It is clear that these terms are *egocentric* terms; they involve the specification of the position of the sound in relation to the observer's own body.

(Evans. 1982, p. 384)<sup>14</sup>

It is a small step from here to the claim that a newly sighted subject would immediately recognize a given visually represented shape whose spatial properties are structured in the very same egocentric frame of reference structuring the subject's tactile recognitional competence with the shape. On these grounds Evans answers in the affirmative to Molyneux's question.

There is a moral here to be carried over to our present question. To account for our perceptual sensitivity to the identity of the body interoceptively and exteroceptively perceived as the spatial hypothesis suggests, it won't be enough to show that there is plenty of spatial content in both interoceptive and exteroceptive perception. It must also be the case that there is a single frame of reference in which this spatial content is structured through the two sets of senses. (To compare, imagine a side-by-side comparison of a standard street map of Paris and unlabelled architectural drawings of the internal structures of the Eiffel tower on the other. Both depict detailed spatial relations in an

overlapping region of space. But differences in how they represent that space would make it impossible to recognize the overlap without appeal to independent “bridging” information between the two depictions.) If interoception and exteroception encode spatial information in distinct frames of reference, then it’s not clear how our processing systems could reliably exploit recognized spatial coincidences between them to resolve interoceptive-exteroceptive binding tasks.

The question now is, do we have any reason to think that there is a single frame of reference structuring spatial contents across interoception and exteroception? The spatial hypothesis presupposes a “yes” answer to this question. An initial reason to suspect otherwise is that this exteroceptive egocentric frame of reference is *centred on* and *radiates outwards* from the perceiver’s body: ahead (of me), behind (me), (to my) left, (to my) right. These egocentric terms simply don’t seem to capture the felt locations of interoceptive percepts that seem to be located *inside* the very bodily object at the centre of that reference frame. So at least on the face of it, the egocentric exteroceptive reference frame doesn’t seem descriptively adequate for capturing the spatial contents of interoceptive perception. Here are two more reasons to think that interoception and exteroception do not share a reference frame.

### **(i) The Argument From Action**

Evans, we have seen, appealed to sameness of “behavioural space” to answer an analogue of the foregoing question in the purely exteroceptive case. A shared significance for action of the spatial contents of perception across the different exteroceptive modalities invited the postulation of a shared exteroceptive spatial frame of reference. And as we saw, the fact that this significance is typified by the utility of spatially representing one’s environment *relative to one’s body*, in a way can be used to directly inform trajectory-computations for bodily movements required to act on (reach, touch, duck, sit on, pull, kick) objects nearby, gave us a reason to posit a shared *egocentric* spatial frame of reference across the exteroceptive modalities. ■

It’s much less obvious that this strategy will be workable in the mixed interoceptive-exteroceptive case. That’s because we don’t normally act in the space accessed through the interoceptive sense modalities in anything like the ways in which we act either in exteroceptively perceived extra-bodily space or on the exteroceptively perceived external surfaces of the bodily object itself. That is, we don’t normally reach, or swerve to avoid, or approach, or pick up, or push things around inside our interoceptively perceived body spaces—all actions of a kind that might have helped to plot shared points on Evans’s single frame of reference. Of course, this difference isn’t a reason to deny that the spatial contents of interoception are significant for action at all. We clearly act *with* our bodies all the time, and this is plausibly facilitated by

interoceptively gained spatial information about that body, at least in the ordinary case.<sup>19</sup> But the difference is reason enough to resist a simple extension of Evans's argument for the shared exteroceptive frame of reference to the interoceptive senses too.

Added to this negative observation we also have positive reason to expect that the significance of spatial perceptual contents for action will differ between interoception and exteroception, following from how the two sets of senses were distinguished. Interoceptive perception was defined as *single-object dedicated* perception, by which a subject can perceive a single object only (her own body); exteroceptive perception was defined as *multiple-object* perception, in which there is no in-principle constraint on the number of objects a subject can possibly perceive. The suggestion now is that this basic difference between interoception and exteroception issues in a downstream difference in the typical "tracking requirements" for acting directly on their basis. Given its multiple-object nature, acting directly on the basis of exteroceptive perception will typically require such capacities as the ability to identify the perceived target as figure from ground, distinguish it from its neighbours, track it over time, and perhaps later reidentify it in the course of sustained action planning and execution. These are the "tracking requirements" for acting directly on the basis of exteroception, and they plausibly contribute centrally to the significance for action of representing the spatial aspects of an exteroceptively perceived scene.

By contrast, the single-object dedicated nature of interoceptive perception rules such tracking requirements out of relevance when acting directly on the basis of interoception. To the extent that we act on the basis of interoceptive perception alone—say, flexing a muscle in response to feeling a cramp—there is no question of having to locate the body part in question, or of picking it out as figure from ground before acting, as we might do with a rattle heard in the dark. We therefore have positive reason to expect there to be a difference in the significance for action of the spatial contents of the two kinds of perception, just by reflection on the epistemic implications of the foundational difference between them.

Here, then, is a summary of the *argument from action*. (i) The foundational single-/multiple-object difference between interoceptive and exteroceptive perception generates a secondary difference in the typical tracking requirements for acting directly on their basis. (ii) This difference in turn predicts a difference in the typical significance of spatial contents for action between the two forms of perception. So (iii) if perceptual frames of reference correspond to typical significance of spatial contents for action as the Evansian response to Molyneux's question suggests, then this gives us reason to posit distinct frames of reference for interoceptive and exteroceptive spatial contents respectively.

## (ii) The Argument From Non-perspectival Structure

## 16 The Inside-Out Binding Problem

The foundational difference between the two sets of senses also provides the beginnings of a second reason to reject the claim that there is a single frame of reference shared between them. Exteroceptive perception, we have seen, is multiple-object perception. This implies that whatever the contingent features of the individual exteroceptive sensory systems, they must service the possibility of perceiving multiple different objects—however they function, it must be possible at least in principle to “swap out” one object of perception for another. Now this requirement in turn imposes what we might think of as a *perspectival* constraint on exteroceptive perception—that is, that the object perceived must be non-identical to the parts of the body most directly involved in perceiving it, so as to allow the latter to remain and the object to differ. How this constraint is met is most obvious in the case of vision and hearing, where there is typically some distance between the relevant parts of the body and the object perceived. But even in the case of touch, the fact that skin operates by tactile registration of *contact* between the surface of the body and the perceived object provides a way of satisfying this perspectival requirement.

All of this stands in contrast with interoceptive perception, whose single-object dedication exempts the need for a corresponding perspectival requirement. Rather, the parts of the body involved in interoception are themselves parts of the one and only object that is always and only its own interoceptive target. Unlike exteroceptive perception, when I feel a flutter in my stomach or a bend in my knee there need be no distinction between the parts of the body most directly involved in the perceiving and the parts of the body being perceived. The relevance of this non-perspectival structure for present purposes is that, as Bermúdez writes of the proprioceptive case, “there is no privileged part of the body that counts as *me*” (2005, p. 309); given the mandatory non-perspectival nature of these sensory systems there is no non-arbitrary point, or set of points, or axes, on the body, that could be used to determine an egocentric frame of reference for the interoceptive modalities. No ego, this argument goes, so no egocentric frame of reference.

These two arguments leave us with what Lucy O’Brien has elsewhere described as a “poncho-shaped” demarcation between two perceived regions of space at a time: “the space centered around the body of the subject with a hole in the middle” (2007, p. 192). Of course, the idea isn’t that the space in the middle of the poncho is out of perceptual range, or that it is ontologically distinct from the space that surrounds it, or even that it couldn’t be exteroceptively perceived—either by moving out of the way or by forcibly revealing one’s bodily insides. The suggestion, rather, is that with these arguments we now have reason to think that in episodes of interoceptive perception, the space perceived inside the bodily object is structured in a distinct frame of reference from the egocentric reference frame structuring our exteroceptive perceptions.<sup>18</sup> If that’s right, then it is hard to see how the interoceptive-exteroceptive binding problem could be solved by prior perceptual sensitivity to represented spatial overlap across the two sets of

senses at a given time as the spatial hypothesis proposes. Rather, if interoception and exteroception do not share a common spatial reference frame, then our non-sceptical interoceptive-exteroceptive binding problem becomes, in part, a question about how these distinct spatial frames of reference are subpersonally reconciled.<sup>19</sup>

#### 4. Represented Temporal Overlap

It might be helpful at this point to remind ourselves where we've got to so far. We began with the twin ideas that (i) interoception is a form of bodily self-perception through which a subject can directly perceive herself *as herself* but not as an object spatially located among others, and (ii) exteroception is a form of bodily self-perception through which a subject can directly perceive herself *as an object spatially located among others* but not under a *de se* mode of presentation. If our *de se* self-conceptions as objects located among others really do emerge as the result of endorsing the contents of bodily self-perception, then in episodes of interoceptive-exteroceptive crossmodal bodily self-perception, our perceptual systems must be sensitive to the identity of the body perceived through both forms of perception at once. How is this possible? In the last section we saw that a seemingly natural answer to this question—that the possibility is underwritten by recognized spatial co-location of interoceptively and exteroceptively perceived body parts at a time—runs into trouble. This might seem to leave us in the following disquieting position described by Brian O'Shaughnessy:

Instead of the Cartesian consciousness alone with its thoughts, attempting to build a bridge to the beyond, it might seem that we have put in its place the animal alone with its body and its immediately given posture and movements, attempting to build a cognitive bridge, if not to another ontological realm, to the vast remainder of the physical realm.

(1989, p. 54)

In this section I want to suggest a different way out of this position, one that focusses not on spatial but on certain *temporal* aspects of crossmodal bodily self-perception.

Bodily self-perception characteristically involves a large amount of high-resolution temporal information about movements in, on, and with the body through both interoception and exteroception. I am currently interoceptively aware of my right leg's relentless jiggling, and of the quick jerks of my typing fingers that are periodically suspended and replaced by controlled stillness in my fingers and comparatively long, slow, and smooth movements of my arm as it delivers spoonfuls of soup to my mouth. During the same stretch of time, I receive this same movement-information about my fingers and arm in the lower periphery of my visual field, and hear my bouncing knee knocking

## 16 The Inside-Out Binding Problem

against the underside of the table. Such platitudinous observations—I invite you to try some for yourself—grant initial respectability to the suggestion that there is plenty of temporal content to crossmodal interoceptive-exteroceptive bodily self-perception that might serve as the basis for an alternative response to our how-possible question from §1.

More precisely, the alternative suggestion is this:

*Temporal hypothesis.* Personal-level recognition of identity of body(-part) perceived both interoceptively and exteroceptively is at least partly made possible by subpersonal binding of cues from both forms of perception grounded on recognised temporal convergence in movement of body(-parts) perceived both ways at once.

Just as the spatial hypothesis was not atemporal, the temporal hypothesis is not aspatial. I repeat the spatial hypothesis here for comparison:

*Spatial hypothesis.* Personal-level recognition of identity of body(-part) perceived both interoceptively and exteroceptively is made possible by subpersonal binding of cues from both forms of perception grounded on recognised co-location at a time of body(-part) perceived both ways at once.

The temporal hypothesis says that our perceptual sensitivity to the identity of a given body or body part perceived in both ways at once exploits recognized temporal coincidences *in bodily movements*—that is, diachronic change in spatial location. The distinctive suggestion here (as contrasted with the spatial hypothesis) is that it is these recognized temporal coincidences (rather than recognized spatial coincidences) that partly account for the possibility of resolving our interoceptive-exteroceptive binding problems.

Is there a principled reason to think that the temporal hypothesis is true? I think there is, and it is a version of a reason we have seen before—namely, a temporal analogue of Evans’s behavioural argument for a crossmodally shared exteroceptive frame of reference. That argument, recall, moved from claiming a shared crossmodal significance for action of the spatial contents of exteroceptive perception (e.g., there is a common role in action planning/execution for the spatial information gained by perceiving the rattle regardless of whether I see, feel, or hear it in the dark) to positing a shared exteroceptive crossmodal egocentric frame of reference (e.g., to serve this role, in each case, the rattle is perceived as being located in a given position *relative to me*). From here it is no great stretch to expect—as Evans did in his response to Molyneux’s question—that all things equal, spatial coincidences will be readily recognized in crossmodal exteroceptive perception.

Consider now the significance for action of the *temporal* contents of our perceptions. If I am to act directly on the basis of a perceived rattle, it is not enough that I perceive it without perceiving it as being temporally located (when should I begin my grabbing action?). Neither would it help if I were to

perceive it as being temporally located *at-some-time-or-other*, or at a determinate time relative to some arbitrary non-present temporal point—the moment of my death, say, or the big bang. Rather, if my perception of the rattle is to directly inform my capacity to act on it, I must perceive it as being temporally located *now*, or *simultaneous-with-my-perceptual-experience-of-it*.<sup>20</sup> This is so whether I perceive it by vision, audition, or touch. And as in the spatial case, it is a small step from here to the expectation that these perceived temporal coincidences will be readily recognizable across the exteroceptive modalities.

Now, one of the obstacles to extending Evans's egocentric exteroceptive frame of reference to the spatial contents of interoception was that the significance for action of the spatial contents of interoception did not seem to be the same as the agentive significance of the spatial contents of exteroception. So whereas we had an action-based reason to extend the shared reference frame across audition, touch, and vision, we apparently lacked the same reason to extend it to the interoceptive modalities too. I have just suggested that we have reason to think that there's a single temporal frame of reference structuring the temporal contents of exteroceptive perception. Does an analogue of this obstacle prohibit extending this shared exteroceptive temporal frame of reference to the interoceptive modalities too?

No, or at least not obviously. Whether intentionally acting directly on the basis of interoception or exteroception, an actor must make choices about when to begin and end an action—I might opt to start tensing my calf muscle *now* and stop *now*, say, or perhaps to vacillate for a spell between contractions and relaxations of the muscle to relieve a felt cramp. At least in cases of intentional action it is reasonable to require that the agent must minimally be able to keep track of the temporal contours of her action, a tracking condition that applies no less in the interoceptive realm (perceiving-and-acting-on a cramp) than in exteroception (perceiving-and-acting-on a rattle); after all, even if interoceptive modalities are single-object dedicated, they are not dedicated to a single *time*, which must then be kept track of in the normal way.

We have, it seems, a principled reason to expect there to be shared agentive significance to the temporal contents of perception across both interoception and exteroception, and corresponding reason to expect those contents to be structured in a common temporal frame of reference. This is the temporal analogue of Evans's behavioural argument for a single crossmodal exteroceptive spatial frame of reference; there is only one "behavioural *time*," so only one temporal frame of reference. We therefore find ourselves without the same obstacle as in the case of the spatial hypothesis to thinking that the temporal hypothesis is true.

## 5. Empirical Predictions

## 16 The Inside-Out Binding Problem

The question of how the interoceptive-exteroceptive binding task is resolved is an empirical question with a contingent answer. This answer will plausibly be different in our case from other animals, and even intraspecies there will surely be differences between typical and atypical cases. I have argued that there are some broadly *a priori* considerations that appear to rule out the spatial hypothesis (§§2–3) and that appear to favour the temporal hypothesis (§4)—at least as an answer to the how-possible question from §1 posed for perceivers like us, whose sensory modalities can be divided into those that are and aren't dedicated to a single object (one's own body). These philosophical considerations throw up a number of empirical predictions.

The first prediction is that if it's right that the spatial hypothesis is false and the temporal hypothesis is true, then we should expect our crossmodal bodily perception to break down in certain characteristic ways. On the one hand, where there are conflicting candidate resolutions to an interoceptive-exteroceptive binding task in which there is experienced spatial convergence and experienced temporal discrepancy across the two sets of sensory systems, we should expect the binding task to be resolved in alignment with the temporal discrepancy rather than the spatial convergence. So we should expect it to be possible to induce “false-negative” experiences as of numerically distinct body(-parts) in cases in which there is experienced spatial convergence across interoceptive and exteroceptive modalities but experienced temporal asynchrony. On the other hand, where there are conflicting candidate resolutions to an interoceptive-exteroceptive binding task in which there is experienced spatial discrepancy and experienced temporal convergence, we should expect the binding task to be resolved in alignment with the experienced temporal convergence rather than the experienced spatial discrepancy. So we should expect it to be possible to induce “false-positive” experiences as of numerically identical body(-parts) where there is experienced spatial distinctness across interoceptive and exteroceptive modalities but experienced temporal synchrony.

It's not clear how the first possibility might be tested for.<sup>21</sup> A well-known experimental paradigm that might be taken to demonstrate the second possibility, however, is the rubber hand illusion (RHI), at least in some of its variants. In the classic RHI one of the subject's hands is hidden from view and an artificial rubber hand is placed directly in their line of vision. The orientation and distance of the rubber hand conform to ordinary architectural constraints of a biological hand. Experimenters have found that when the rubber hand and the real hand are stroked *synchronously* with a paintbrush (the illusion does not work when the strokes are sufficiently asynchronous),<sup>22</sup> many subjects undergo an experience of their own stroked hand as being at the visually perceived location of the rubber hand. When asked to reach towards their stroked hand with their free hand, subjects undergoing the illusion tend to direct their reach towards the rubber hand. Many also report feeling as if the rubber hand *is* their own hand. These results have been



replicated many times.<sup>23</sup> Interestingly, the RHI has also been shown to arise in cases in which the induced intermodal conflict is between vision and the interoceptive sensory detection of one's own heart rate, by the use of pulsing colour signals in the rubber hand (Suzuki, Garfinkel, Critchley, & Seth, 2013), and between vision and interoceptive kinaesthetic perception of finger movement (Kalckert & Ehrsson, 2017). In all of these cases the RHI is plausibly understood as a primarily spatial illusion; the subject misperceives her own hand as having a locational property other than it in fact has. Specifically, the subject perceptually mislocates her interoceptively felt hand as being at the egocentrically specified location of the seen hand.<sup>24</sup>

The important finding for our purposes is the fact that these interoceptive-exteroceptive binding tasks are resolved by binding together two *in fact spatially separate* objects. This is *pro tanto* evidence that experienced spatial overlap is not the most influential binding principle in the resolution of this interoceptive-exteroceptive binding task; the subjects' integrative processing of interoceptive and exteroceptive cues has been otherwise influenced in such a way as to induce an erroneous experience of these distinct objects as being co-located. We might conclude either that the spatial binding principle is not effective at all in these cases, or that it is, but is trumped by a more dominant alternative principle that overrides it to generate the spatial illusion of co-located-ness in the first place. So understood, the RHI seems to be an example of just the kind of "false-positive" illusion of body-part identity raised earlier as a prediction of the findings of the last two sections.

A further, albeit speculative, neurological prediction is that if the spatial hypothesis is false then we might expect at least partially distinct neural systems to be involved in the processing of spatial and binding information respectively in interoceptive-exteroceptive crossmodal perceptual processing, so as to allow for information about the binding of an object to be dissociable from information about perceived spatial properties. (There need be no corresponding expectation, if the temporal hypothesis is true, that there is a single neural mechanism carrying information about both temporal and identity perception; either set-up is consistent with the truth of the temporal hypothesis.) The neural mechanisms involved in interoceptive-exteroceptive binding are not presently well understood.<sup>25</sup> However, in the case of exteroceptive binding, Calvert, Brammer, and Iversen (1998) have argued that there are distinct mechanisms underlying location (where?) judgments and identity (what?) judgments in the case of audiovisual integration, a division of processing labour that mirrors the visual processing of location information and identity information in two separate neural pathways in the visual system—the dorsal and ventral streams respectively (p. 248). That there are multiple dissociable pathways in these cases might be taken as suggestive that similar structures will be present in the interoceptive-exteroceptive case too. This prediction is further supported by emerging evidence that the experience of "body ownership" on the one hand and the experience of bodily location

on the other (both tested in paradigms involving interoceptive-exteroceptive integration tasks) occupy distinct neural substrates: the premotor cortex for bodily ownership, and the temporo-parietal junction for self-location (Serino et al., 2013).

Finally, if the spatial hypothesis is false and our capacity to reconcile the interoceptive and exteroceptive frames of reference really is developmentally grounded in a prior capacity to resolve the interoceptive-exteroceptive binding problem, then we might expect this ordering to show up in developmental findings from infants and young children. It is well established that infants are perceptually sensitive to certain temporal relations (synchrony, duration, rate, and rhythm) across different perceptual modalities from a very early age.<sup>26</sup> Indeed, it has even been suggested that awareness of temporal synchrony is involved in crossmodal integration *before* birth; as David Lewkowicz writes,

given that the human fetus is sensitive and responsive to stimulation in different sensory modalities, it is not difficult to imagine how the co-occurrence of stimulation in different modalities might contribute to the early appearance of postnatal responsiveness to temporally contiguous heteromodal inputs.

(2000, p. 294)

The early development of capacities for the detection of crossmodal temporal synchrony is, of course, not decisive, but it is suggestive.

The idea that the temporal contents of perception play a foundational role in crossmodal integration is by no means new to the empirical literatures on crossmodal perceptual processing; as Lawrence Marks observed in the 1970s, “[I]f there is any attribute that truly deserves to be called a common sensible, any attribute of objects or events that really can manifest itself through *all* of the senses, it is time” (1978, p. 32, original emphasis).<sup>27</sup> The suggestion of this chapter has been that we put this received idea to work in explaining an even older idea: that there are certain demands on our self-conscious thought that require the capacity to conceive of ourselves in a *de se* way as objects spatially located among others. Kant argued that we encounter our mental lives ordered in time through empirical self-consciousness, and for that reason we must be aware of our bearing spatial relations to objects outside us. The suggestion here has been that our capacity to meet this requirement is itself grounded in the temporal contents of perception that cut across the interoceptive and exteroceptive senses. This suggestion too is clearly Kantian in spirit: space might be the form of outer sense, but it is time that is the universal formal condition of *all* appearances. (A34/B50–51).

## 6. Conclusion

## 16 The Inside-Out Binding Problem

The requirement that we are able to conceive of ourselves *as ourselves* and as objects spatially located among others recurs as a thematic drumbeat in certain strands of post-Kantian philosophy of mind. An initially plausible suggestion is that we develop these self-conceptions by endorsing the contents of bodily self-perception. In §1 I raised a *how-possible* challenge for this suggestion, taking off from the idea that different forms of bodily self-perception provide unmediated grounds for very different kinds of self-directed thought. Interoceptive perception directly sustains self-directed thought under a *de se* mode of presentation, but not as an object spatially located among others, whereas exteroceptive self-perception sustains thought about oneself as an object bearing spatial relations to others, but not thought of a *de se* kind. How is it possible that in episodes of interoceptive-exteroceptive crossmodal bodily self-perception we are able to recognize the identity of the body (and its parts) perceived both ways at once, given the apparent absence of substantial property overlap between them?

There could, of course, be more than one answer to this challenge. But we now seem to be in a position to dissolve it in at least one way. It is possible because there *is* substantial property overlap between the two sets of senses: even if it's right that we don't perceive space in a way that crosses the bodily boundaries, we have reason to think that the *temporal* contents of perception cut across the interoceptive-exteroceptive divide.<sup>28</sup>

## Notes

## References

- Allison, H. (1983). *Kant's transcendental idealism: An interpretation and defense*. New Haven, CT and London: Yale University Press.
- Aymerich-Franch, Laura, Damian Petit, Gowrishankar Ganesh, and Abderrahmane Kheddar, 2016. 'The second me: Seeing the real body during humanoid robot embodiment produces an illusion of bi-location', in *Consciousness and Cognition* (46), pp. 99-109.
- Armel, K. C., & Ramachandran, V. S. (2003). Projecting sensations to external objects: Evidence from skin conductance response. *Proceedings of the Royal Society B: Biological Sciences*, 270(1523), 1499–1506.
- Bermúdez, J. L. (1995). Ecological perception and the notion of a nonconceptual point of view. In N. Eilan, J. L. Bermúdez, & T. Marcel (Eds.), *The body and the self*. Cambridge, MA: MIT Press.
- Bermúdez, J. L. (1998). *The paradox of self-consciousness*. Cambridge, MA: MIT Press.
- Bermúdez, J. L. (2005). The phenomenology of bodily awareness. In D. W. Smith, & A. L. Thomasson (Eds.), *Phenomenology and philosophy of mind*. Oxford: Oxford University Press.

## 16 The Inside-Out Binding Problem

- Bermúdez, J. L. (2017). *Understanding “I”: Language and thought*. Oxford: Oxford University Press.
- Blanke, O. (2012). Multisensory brain mechanisms of bodily self-consciousness. *Nature Reviews Neuroscience*, 13, 556–571.
- Botvinick, M., & Jonathan, C. (1998). Rubber hands “feel” touch that eyes see. *Nature*, 391(6669), 756.
- Brewer, B. (1995). Bodily awareness and the self. In N. Eilan, J. L. Bermúdez, & T. Marcel (Eds.), *The body and the self*. Cambridge, MA: MIT Press.
- Campbell, J. (1994). *Past, space and self*. Cambridge, MA: MIT Press.
- Calvert, Gemma A., Michael J. Brammer, and Susan D. Iversen, 1998. 'Crossmodal identification', in *Trends in Cognitive Sciences*, 2:7.
- Campbell, J. (2002). *Reference and consciousness*. Oxford: Oxford University Press.
- Cassam, Q. (1993). Inner sense, body sense, and Kant’s “Refutation of Idealism”. *European Journal of Philosophy*, 1(2), 111–127.
- Cassam, Q. (1997). *Self and world*. Oxford: Oxford University Press.
- Cassam, Q. (2007). *The possibility of knowledge*. Oxford: Oxford University Press.
- Cole, J. (2016). *Losing touch: A man without his body*. Oxford: Oxford University Press.
- Constantini, M., Robinson, J., Migliorati, D., Donno, B., Ferri, F., & Northoff, G. (2016). Temporal limits on rubber hand illusion reflect individuals’ temporal resolution in multisensory perception. *Cognition*, 157, 39–48.
- Erhsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science*, 317, 1048.
- Ehrsson, H. H., Holmes, N. P., & Passingham, R. E. (2005). Touching a rubber hand: Feeling of body ownership is associated with activity in multisensory brain areas. *Neuroscience*, 25(45), 10564–10573.
- Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.
- Evans, G. (1985). Molyneux’s question. In his *Collected Papers*. Oxford: Oxford University Press.
- Furnaletto, Tizano, Cesare Bertone and Cristina Beccio, (2013). 'The bilocated mind: new perspectives on self-localization and self-identification' in *Frontiers in Human Neuroscience*, (7)
- Guterstam, A., & Erhsson, H. H. (2012). Disowning one’s seen real body during an out-of-body illusion. *Consciousness and Cognition*, 21, 2.
- Guterstam, A., Gentile, G., & Ehrsson, H. H. (2013). The invisible hand illusion: Multisensory integration leads to the embodiment of a discrete volume of empty space. *Journal of Cognitive Neuroscience*, 25, 7.
- Heydrich, L., Aspell, J. E., Marillier, G., Lavanchy, T., Herbelin, B., & Blanke, O. (2018). Cardio-visual full body illusion alters bodily self-consciousness and tactile processing in somatosensory cortex. *Scientific Reports*, 8, 9230.
- Kalckert, A., & Ehrsson, H. (2017). The onset time of the ownership sensation in the moving rubber hand illusion. *Frontiers in Psychology*, 8, 344.
- Kant, I. (1781/1787/1998). *The critique of pure reason* (P. Guyer & A. Wood, Trans.). Cambridge: Cambridge University Press.

## 16 The Inside-Out Binding Problem

- Kassam, R., & Alexandre, F. (2015). Integration of exteroceptive and interoceptive information within the hippocampus: A computational study. *Frontiers in Systems Neuroscience*, 9, 87.
- Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review*, 88(4), 513–542.
- Lewkowicz, David J., 2000. 'The development of intersensory temporal perception: an epigenetic systems/limitations view', in *Psychological Bulletin*, 126:2.
- Locke, J. (1694). *An essay concerning humane understanding*. London: Printed by Eliz. Holt, for Thomas Basset.
- Marks, Lawrence, 1978. *The Unity of the Senses: inerrelations among the modalities*. Academic Press Series in Cognition and Perception. London: Academic Press.
- Martin, M. G. F. (1997). Self-observation. *European Journal of Philosophy*, 5(2), 119–140.
- Merleau-Ponty, M. (1989). *Phenomenology of perception*. London: Routledge.
- Nagel, T. (1986). *The view from nowhere*. New York, NY: Oxford University Press.
- O'Brien, L. (2007). *Self-knowing agents*. Oxford: Oxford University Press.
- O'Shaughnessy, B. (1989). The sense of touch. *Australasian Journal of Philosophy*, 67(1), 37–58.
- O'Shaughnessy, B. (2008). *The will: A dual aspect theory*. Cambridge: Cambridge University Press.
- Perry, J. (1979). The problem of the essential indexical. *Noûs*, 13, 3–21.
- Prosser, S., & Recanati, F. (2012). *Immunity to error through misidentification: New essays*. Cambridge, UK: Cambridge University Press.
- Ramachandran, V. S., & Blakeslee, S. (1998). *Phantoms in the brain: Probing the mysteries of the human mind*. New York, NY: William Morrow.
- Rosén, B., Ehrsson, H. H., Antfolk, C., Cipriani, C., Sebelius, F., & Lundborg, G. (2009). Referral of sensation to an advanced humanoid robotic hand prosthesis. *Scandinavian Journal of Plastic and Reconstructive Surgery and Hand Surgery*, 43(5), 260–266.
- Salje, L. (2017). Crossed wires about crossed wires: Somatosensation and immunity to error through misidentification. *Dialectica*, 71(1), 35–56.
- Serino, A., Alsmith, A., Constantini, M., Mandrigin, A., Tajadura-Jimenez, A., & Lopez, C. (2013). Bodily ownership and self-location: Components of bodily self-consciousness?. *Consciousness and Cognition*, 22(4), 1239–1252.
- Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. *Annals of the New York Academy of Sciences*, 1296, 31–49.
- Strawson, P. F. (1966/1990). *Bounds of sense: An essay on Kant's critique of pure reason*. New York, NY: Routledge.
- Suzuki, K., Garfinkel, S. N., Critchley, H. D., & Seth, A. K. (2013). Multisensory integration across exteroceptive and interoceptive domains modulates self-experience in the rubber-hand illusion. *Neuropsychologia*, 51, 2909–2917.

## 16 The Inside-Out Binding Problem

Treisman, A. (2003). Consciousness and perceptual binding. In A. Cleeremans (Ed.), *The unity of consciousness: Binding, integration and dissociation*. Oxford: Oxford University Press.

---

Or, at least, arguably argues; see Henry Allison (1983) and Cassam (1993, 1997) for defences of this reading.

Strawson (1966), Nagel (1986), Evans (1982), Campbell (1994), Cassam (1997), and Bermúdez (1995, 2017).

On one way of counting objects, not relevant to present purposes, the body's parts comprise multiple interoceptively perceivable objects; for a defence of the notion of single-object bodily perception see M.G.F. Martin (1997).

These terms are borrowed from psychology and the neurological and biological sciences, but there is no single accepted way of defining these terms in those literatures; the suggested epistemic distinction is the one relevant to the philosophical issue raised here.

The most theoretically useful formulation of IEM is a matter of ongoing debate (see, e.g., chapters in Prosser & Recanati, 2012); the details of those debates are not relevant to this chapter. The standard challenge to the claim that interoception is a source of judgments with IEM comes from "crossed wire cases"; I have argued against that challenge elsewhere (Salje, 2017), and will simply assume the IEM of interoceptive judgments (and non-IEM of exteroceptive judgments) here.

This is distinct from and compatible with the claim that self-locating judgments, such as "I am in front of a burning tree," made on visual grounds are IEM relative to uses of the first-person concept; the capacity to form such *de se* self-locating judgments plausibly interacts in interesting ways with the capacity to solve the binding problem raised in this chapter, but discussion of their connection falls outside the scope of this chapter.

Thanks to an anonymous reviewer for pointing out that this is not the only way of forming such a self-conception, even accepting both suppositions. Strictly speaking if one was presented to oneself in interoception as both *oneself* and *the unique F*, and in exteroception as both *the unique F* and as *an object spatially located among others*, then jointly endorsing the contents of these independent experiences could generate a complex *de se* conception of oneself as an object located among others without reliance on crossmodal interoceptive-exteroceptive self-perception. Given, however, that it is independently plausible that we have crossmodal interoceptive-exteroceptive perceptions of ourselves, the simpler suggestion that we develop these self-conceptions by endorsing the contents of these crossmodal perceptions deserves serious consideration. In any case, decoding "*the unique F*" would itself require a solution to some form of the interoceptive-exteroceptive binding problem raised here.

For a book-length discussion, including various applications of this framework, see Cassam (2007).

This is the same level of explanation that concerned Evans in arguments considered in later sections of this chapter: "It is perfectly consistent with the *sense* that I have assigned to this [egocentric] vocabulary that its terms should *refer* to points in a public three-dimensional space" (1982, p. 157, original emphasis).

## 16 The Inside-Out Binding Problem

---

- <sup>10</sup> See Spence's chapter in this volume for a case against the spatial rule as the universal binding principle in exteroceptive crossmodal integration; see also Spence (2013).
- <sup>11</sup> Smell and taste carry less spatial information than vision, touch, and audition; I will bracket them for the rest of this chapter. Likewise for non-spatial interoceptive senses, like the sensory system for the detection of fatigue.
- <sup>12</sup> "Suppose a Man born blind, and now adult, and taught by his touch to distinguish between a Cube, and a Sphere of the same metal, and nighly of the same bigness, so as to tell, when he felt one and t'other; which is the Cube, which the Sphere. Suppose then the Cube and Sphere placed on a Table, and the Blind Man to be made to see. Quaere, Whether by his sight, before he touch'd them, he could now distinguish, and tell, which is the Globe, which the Cube" (Locke, 1694; Book II, IX, p. 8).
- <sup>13</sup> See Evans (1982, p. 156).
- <sup>14</sup> These egocentric terms need not be understood as providing a reductive analysis of the spatial contents of perception; Evans claims only that "when I speak of information 'specifying a position in egocentric space,' I am talking [. . .] of a special kind of information about space—information whose content is specifiable in an egocentric vocabulary" (1982, p. 157).
- <sup>15</sup> See Bermúdez (2005, §5) for a detailed account of what he calls *object-relative spatial coding*, and references contained therein for corroborating empirical results.
- <sup>16</sup> Things will clearly be different in extraordinary cases in which such spatial information is unavailable, as in the case of Ian Waterman, a subject who lost all sense of touch or interoceptive sense of position or movement from the neck down as a result of extreme nerve-damage but has since learned how to guide bodily movement with vision (Cole, 2016).
- <sup>17</sup> A slight complication to this picture is that we *do* sometimes seem to act directly on the basis of interoception in a way that parallels the way that I act on an auditory perception of a rattle heard in the dark—for example, when I move my hand to rub my cramping muscle. However, actions of this kind rely on spatial information coming from both interoceptive and exteroceptive modalities, so are not examples of acting on the basis of interoceptive perception alone. I could not perform my rubbing action, for instance, if I did not locate the muscle "from the outside" by tactile registration of resistance to touch. Given that the interoceptive-exteroceptive binding question is not a sceptical question, our ability to act in ways that depend on its resolution is to be expected (however that initial binding question is resolved).
- <sup>18</sup> The most worked-out account of these distinct frames of reference I have come across is in Bermúdez (2005).
- <sup>19</sup> What about the suggestion that a binding principle of spatial co-location doesn't require spatial information to be encoded in the *same* frame of reference, but only *easily translatable* frames of reference? Assuming that "easy translation" here means that it is possible to reliably map points represented in one frame of reference onto another, this suggestion gets us no further than where we began: our how-possible question is not a sceptical question, so it is assumed throughout that it is *possible* to recognize the spatial co-location of an interoceptively and exteroceptively perceived body part; our question is *how* it is possible. The "translatability" suggestion does not offer a solution to this question, just another way of posing it.

- <sup>20</sup> The argument is neutral on whether the temporal contents of experience are best construed in A-theoretic or B-theoretic terms.
- <sup>21</sup> Though some *heautoscopie* illusions might be relevant here, “out-of-body” crossmodal illusions in which one experiences oneself to be multiply located at a time (see, e.g., [Erhsson, 2007](#); [Guterstam & Erhsson, 2012](#), Furlanetto et al., 2013, Aymerich-Franch et al., 2016). While these demonstrate that self-directed spatial location judgments based on crossmodal experience can be dissociated from self-identification judgments based on crossmodal experience, the role in these illusions of temporal perception is not sufficiently clear for these to constitute decisive examples of the false-negative possibility raised here.
- <sup>22</sup> See [Constantini et al. \(2016\)](#) for evidence that how much asynchrony is tolerated in the RHI corresponds to individual temporal resolution in (exteroceptive) crossmodal integration, but is typically very short (500–1000 ms).
- <sup>23</sup> The original experiment is from [Botvinick and Cohen \(1998\)](#). The cases of intermodal conflict that have been found to produce these kinds of illusions need not involve the typically dominating sense of vision; see, for example, [Ramachandran and Blakeslee \(1998\)](#) for a version based on “inside” and “outside” senses of touch.
- <sup>24</sup> Or in some cases, as being at an egocentrically specified location in a discrete volume of empty space ([Guterstam et al., 2013](#)).
- <sup>25</sup> [Heydrich et al. \(2018\)](#); though see also [Kassam and Alexandre \(2015\)](#).
- <sup>26</sup> See Lewkowicz (2000) and references contained therein.
- <sup>27</sup> Cited in Lewkowicz (2000, p. 285).
- <sup>28</sup> Thanks very much to Lucy O'Brien, Max Jones, Daniel Morgan, Ali Boyle, Harry Farmer, and Alisa Mandrigin for comments and discussion on earlier drafts of this chapter, and to an anonymous referee for this volume.