The Zero Point and I

Terry Horgan
Shaun Nichols
University of Arizona

Der jeweilige Aspect hat seine notwendige Beziehung auf den Nullpunkt der Orientierung, auf das absolute “Hier”
Edmund Husserl, Ideen II, p. 127

1. Introduction

One of the most influential ideas on subjectivity from the analytic tradition comes from Hector-Neri Castañeda and John Perry. The idea is that first-person pronouns are sometimes used in an important and distinctive way, viz., to play the role of the first-person “essential indexical.” Perry memorably describes following a trail of sugar around a grocery store before realizing, “I was the shopper I was trying to catch” (Perry 1979, 3). Once Perry realized this, he fixed the leaky bag of sugar. He explains this new behavior by attributing to himself the new belief “I am making a mess”. This explanation wouldn’t work, Perry argues, if the term “I” is replaced by a description or a proper name, including his own name. For Perry can have a belief that he would express by saying “Perry is making a mess” without having a belief he would express by saying “I am making a mess”. A natural extension of this idea that “I” operates as an essential indexical – indeed, it is perhaps already implicit in Castañeda & Perry – is that there a corresponding kind of mental representation which has this function in thought—perhaps a first-person indexical term in the “language of thought” (see, e.g., Rey 1997), perhaps a first-person, essentially indexical, imagistic or pictorial form of self-representation, or perhaps both. With a nod to Castañeda, we will use the notation I* (and me*) for such a first-person, essentially indexical, kind of mental representation.

A second important idea about subjectivity occurs in both the analytic and the continental traditions. In analytic philosophy of mind, it arises in connection with phenomenal consciousness. Phenomenally conscious mental states are typically described as those mental states such that there is something that it is like, for the experiencer, to undergo such a state. The “what it is like” aspect it often called phenomenal character, whereas the “for the experiencer” aspect is what we will here call phenomenal subjectivity. Phenomenal subjectivity, we take it, is essentially the same feature which, in the continental tradition, is sometimes called pre-reflective self-consciousness, or pre-reflective self-awareness.

A third significant idea, concerning certain perspectival aspects of conscious experience, is what Husserl called the experiential zero point (der Nullpunkt). A very similar idea sometimes surfaces in discussions of mental representation in the analytic tradition—for instance, the notion of “scenario content” advanced and deployed in Peacocke (1992 Ch. 3).

In an important respect, the second and third ideas both are sparser than the idea of I*/me* representations: viz., they both refer to aspects of subjectivity that are already present in each conscious experience whether or not that experience deploys any explicit self-representations.

We will focus here on two guiding questions. First: When, and why, are various cognitive tasks faced by humans accomplished in the sparser way, without the use of I*/me* mental representations? Second: When, and why, are certain cognitive tasks instead accomplished in the less sparse way, by deploying I*/me* representations?

We will address each of these questions in two complementary, mutually reinforcing, ways. On one hand we will resort to phenomenological description: giving attentive introspective characterizations
of certain aspects of conscious experience, characterizations that we hope will resonate with you the reader. On the other hand, we will also invoke pertinent aspects, both empirical and theoretical, of recent and contemporary cognitive science. Along the way we will have things to say about how we think these two kinds of inquiry are related to one another, and about other aspects of the two-pronged methodology we here employ.

2. Phenomenal Subjectivity, Non-Representational Self-Presence, and the Zero Point
2.1. Preliminaries

In recent and contemporary philosophy of mind in the analytic tradition, there is growing advocacy for the contention that virtually every aspect of human mentality that is conscious-as-opposed-to-unconscious is also phenomenally conscious—and, moreover, has phenomenal character that is proprietary (i.e., is uniquely specific to the kind of attitude involved—e.g., belief, hope, desire, etc.), is distinctive (i.e., is uniquely specific to the particular intentional content, if any, of the given mental state), and is individuative (i.e., is that in virtue of which the given state counts as being the kind of state it is, and counts as having its particular intentional content, if any). There is also growing advocacy for the contention that virtually all types of phenomenal character are inherently intentional—and, moreover, that this phenomenal intentionality both (i) is the fundamental kind of mental intentionality, and (ii) is intrinsic to phenomenal character rather than being constitutively dependent on any historical or causal-covariational connections between one’s phenomenally conscious mental states/processes and one’s wider physical or social environment. (Externalistic elements might still combine constitutively with phenomenal intentionality to fix certain derivative aspects of mental intentionality—e.g., mental reference to individuals in one’s environment, and/or mental reference to natural-kind properties instantiated in one’s environment.) In slogan form, the position can be characterized this way: 

 Phenomenality is inherent to consciousness, and intentionality is virtually always inherent to phenomenality. 

Although this package of views is still a minority position in analytic philosophy of mind, the position’s popularity seems to be increasing rapidly. One of us (Horgan) has been among the philosophers championing such views in print (often collaboratively with John Tienson and/or George Graham); others include Colin McGinn, David Pitt, Charles Siewert, and Galen Strawson. For our purposes in the present paper, we will assume henceforth that the position, as just sketched, is in essential respects correct. (The continental tradition, as we understand it (largely from secondhand sources), pretty much took for the position for granted. Good for them, we say!)

In analytic philosophy of mind during the past 70 years or so, heavy emphasis often has been placed upon the functional roles played by various kinds of mental states/processes. Indeed, during much of this time period, orthodoxy had it that the only kinds of conscious mental states/processes that possess phenomenal character are sensations, such as tickles, smells, pains, and color-sensations. By contrast, the orthodox view of mental states like conscious occurrent thoughts, for example, was that the essential character of such a state, qua mental, consists of nothing more that its distinctive functional role in cognitive economy, vis-à-vis sensory “inputs,” behavioral “outputs,” and other mental states/processes (many of which likewise have an essence, qua mental, that is purely functional and not at all phenomenal). In Ned Block’s highly influential terminology (Block 1995), mental states like conscious occurrent thoughts were said to be “access conscious” but not phenomenally conscious—where access

1 The reason for the qualifier ‘virtually’ is to leave open the possibility that some conscious mental states might have phenomenal character but lack any intentional content—for instance, perhaps certain states of diffuse anxiety.

consciousness is supposedly entirely a functional-role matter involving, for instance, availability of a mental state for accurate first-person characterization of oneself as currently instantiating that very state. (On more recent versions of this orthodoxy, the constitutive functional roles of states like occurrent conscious thoughts are allegedly “long-armed,” extending outward into a creature’s wider environment; thoughts about water, for instance, refer to H20 partly because H20 is the stuff in the wider environment that tends to trigger such thoughts.)

Our discussion in this paper is predicated upon a thoroughgoing rejection of this long-prevalent orthodoxy in analytic philosophy of mind, and upon acceptance of the contention that phenomenality is inherent to consciousness and intentionality is virtually always inherent to phenomenality. Although this alternative orientation does not deny the importance of the functional roles played by various kinds of conscious mental states/processes, it casts these functional roles in a very different light than did the old orthodoxy. The basic new element is this: the inherent phenomenal character of a particular mental state or process is the experiential categorical basis of the pertinent functional-role aspects of that state/process. Consider, for instance, an occurrent conscious thought. Although part of the typical functional role of such a mental state, insofar as it is conscious-as-opposed-to-unconscious, is indeed that the state is accessible—which means, inter alia, that if one is queried about what one is currently thinking, one will easily be able to answer—the state has this functional-role feature in virtue of its inherent, distinctive, phenomenal character. The occurrent thought’s distinctive phenomenal character, with its inherent intentional content, is immediately given to you experientially; this is what enables you to accurately report that you are now undergoing that very occurrent thought.

2.2. Phenomenal Subjectivity as Non-Representational Self-Presence

Phenomenal subjectivity, as we said in Section 1, is the “for me” aspect of phenomenal consciousness; there can be no occurrent “what-it-is-like-ness” apart from an experiential subject for whom there is something that a token phenomenal state is like. Moreover, in phenomenally conscious experience, the experiencing self is not just instantiating the pertinent phenomenal state/process. Rather, the experiencing self is experientially present within the state/process, and is present in a quite direct and intimate way. Such self-presence, via the fact that the experience’s phenomenal character is its distinctive what-it-is-like-ness for me—is more direct and more intimate than would be the presence in consciousness of an overt representation of the self. It is also more direct and more intimate than the way that currently perceivable objects in one’s ambient environment can be said to be “experientially present,” since this latter form of experiential presence is still mediated via overt mental representations—ones that might aptly be called presentational representations. Instead, such experiential self-presence is an immediate—i.e., un-mediated—involvement of the self in consciousness, as opposed to mediated involvement via the conscious representation of self. Phenomenal subjectivity—the “for-me-ness” of phenomenal consciousness without which one could not instantiate states with phenomenal character (i.e., with “what-it’s-like-ness”), is an experientially direct way in which the self is present in phenomenal consciousness.

---

3 The notion of a presentational representation can be characterized, at least roughly and initially, as follows. An experience presentationally represents a putative object, and/or a putative state of affairs, just in case (i) it is a sensory-perceptual experience, and (ii) it represents such a putative object, and/or such a putative state of affairs, as present in the experiencer’s ambient environment. We insert the word ‘putative’ because it is possible in principle to undergo sensory-perceptual states that are radically non-veridical—e.g., hallucinations that are phenomenologically indistinguishable from veridical sensory-perceptual experiences. Indeed, it is possible in principle for one’s sensory-perceptual experience always to be radically non-veridical—e.g., by being a lifelong envatted brain whose sensory-perceptual states are phenomenologically indistinguishable from those of a normal human being.
As we understand it, phenomenal subjectivity, thus construed as a form of non-representational self-presence in phenomenal consciousness, is essentially the same feature that in the continental tradition is sometimes called “pre-reflective self-consciousness” or “pre-reflective self-awareness.” The following passage from Shaun Gallagher and Dan Zahavi (2010) describes this latter notion in a way that explicitly connects it to phenomenality as typically described in analytic philosophy of mind:

The notion of pre-reflective self-awareness is related to the idea that experiences have a subjective ‘feel’ to them, a certain (phenomenal) quality of ‘what it is like’ or what it ‘feels’ like to have them. As it is usually expressed outside of phenomenological texts, to undergo a conscious experience necessarily means that that there is something it is like for the subject to have that experience (Nagel 1974; Searle 1992). This is obviously true of bodily sensations like pain. But is also the case for perceptual experiences, experiences of desiring, feeling, and thinking. There is something it is like to taste chocolate, and this is different from what it is like to remember what it is like to taste chocolate, or to smell vanilla, or to run, to stand still, to feel envious, nervous, depressed or happy, or to entertain an abstract belief. Yet, at the same time, as I live through these differences, there is something experiential that is, in some sense, the same, namely their distinct first-personal character. All the experiences are characterized by a quality of mineness or for-me-ness, the fact that is I who am having these experiences. All the experiences are given (at least tacitly) as my experiences, as experiences I am undergoing or living through. All of this suggests that the first-person experience presents me with an immediate and non-observational access to myself, and that consequently (phenomenal) consciousness consequently entails a (minimal) form of self-consciousness. To put it differently, unless a mental process is self-conscious there will be nothing it is like to undergo the process, and therefore it cannot be a phenomenally conscious process…. It could consequently be claimed that anybody who denies the for-me-ness of experience simply fails to recognize an essential constitutive aspect of experiences. (p. 3)

Although we find ourselves largely on board with the remarks in this passage, nonetheless we are unhappy with the resonances (in English, anyway) of expressions like ‘pre-reflective self-consciousness’ and ‘pre-reflective self-awareness’. These expressions seem problematic in two respects. First, locutions like ‘self-consciousness’ and ‘self-awareness’ strongly suggest the idea that the self is an intentional object of conscious awareness, and thus that the self is being explicitly represented in conscious awareness. (Presumably this is not intended, since the key idea is that self is experientially present without being explicitly represented; but the resonances are there anyway.) Second, the modifier ‘pre-reflective’ strongly suggests the idea that these states are to be contrasted with mental processes of reflection, such as deliberation or temporally extended discursive thought. (Presumably this too is not intended, since the intended contrast is with explicit representation of self, rather than with deliberation; but again, the resonances are there anyway.) So hereafter we will eschew expressions like ‘pre-reflective self-consciousness’ and ‘pre-reflective self-awareness’; instead we will sometimes deploy our own recommended replacement-terminology—viz., ‘non-representational self-presence’—and we will sometimes just deploy the expression “phenomenal subjectivity”.

2.3. **Zero-Point Aspects of Phenomenal Subjectivity**

An enormously fruitful idea we will draw upon from the continental tradition is the one expressed by Husserl’s expression ‘der Nullpunkt’ (the zero point). As we understand it, this expression denotes the experiential perspectivalness of much human conscious experience—including, in particular, sensory-perceptual consciousness. Visual experience, for instance, represents various objects in one’s ambient environment as located relative to a specific vantage point—i.e., relative to a specific
spatio-temporal “absolute ‘Hier’.” As we will emphasize shortly, consciousness exhibits a variety of different aspects of perspectivalness, what we will call zero-point aspects.

Concerning the zero point of visual-perceptual experience, the following question naturally arises. Is this zero point explicitly represented within such experience? The answer, which should be obvious introspectively, is No. Although the various objects that one now sees are each experienced as occupying specific locations—locations all positioned within a spatio-temporal frame of reference centered upon the “absolute ‘Hier’” from which these objects are presentationally represented by one’s current visual-perceptual experience—that perceptual vantage point is not itself being visually-perceptually represented. (Normally one can see parts of one’s own body, of course; but one cannot see the portion of one’s body, or the location, from which one sees the things that currently are presentationally represented visually.)

And so now the following further question naturally arises. Is the visual-perceptual zero point an aspect of phenomenal subjectivity? The answer, which should be obvious introspectively, is Yes. The self that is present in consciousness directly and without the mediation of a self-representation—the me that is experientially present via the for-me-ness of consciousness—is directly present in experience as spatio-temporally located within the ambient environment. That is, the spatio-temporal location of the visual-perceptual zero point affects or colors the phenomenology. Although much of that ambient environment is indeed represented in one’s current consciousness, and is represented via presentational representations of the visual-perceptual kind, one’s own visual-perceptual zero point is present in the direct way – directly influencing the phenomenology - rather than being explicitly represented. One’s visual-perceptual perspectival locus, then, is a zero-point aspect of phenomenal subjectivity. To see this with a concrete example, consider a sailor in a submarine looking down the length of the control room. Now consider the sailor putting the periscope in front of his eyes. This will cause an abrupt shift not just in the content of what he sees, but in the zero-point perspective from which he sees it. His phenomenology without the periscope is as of the perspective spatially inside the submarine; his phenomenology with the periscope is as of the perspective on top of the submarine. These different aspects of the visual-perspectival zero-point will be phenomenologically salient, but not through any form of explicit representation.

In the subsequent subsections of section 2, we will maintain that attentive phenomenological introspection reveals a variety of zero-point aspects of phenomenal subjectivity, some of which involve features of consciousness other than sensory-perceptual.

2.4. Sensory-Perceptual Experience

The zero-point aspects of sensory-perceptual experience are rich, multifarious, and subtle. One dimension of subtlety is the way zero-point aspects combine with aspects of explicit sensory-perceptual representation. In tactile experience, for instance, it seems that various parts of one’s body are themselves explicitly represented: one can feel them, and one can feel them as instantiating various properties both synchronic (e.g., pressure upon the bottoms of one’s feet) and diachronic (e.g., the fingers of one’s hand closing into a fist). On the other hand, one’s tactile experiences often have externally directed intentional content as well (e.g., the shape and heft of the dumbbell one is now holding). Some aspects of external-world-representing tactile experience seem centered on one or several zero points that coincide with certain specific body parts that themselves are also explicitly represented tactilely; for instance my tactile experience of the dumbbell’s shape seems centered on my two hands, one of which feels (via gripping) the cylindrical shaft and the other of which feels the shape of one of the weight-plates to which the shaft is attached. Other aspects of external-world-representing tactile experience—for instance, my tactile experience of the dumbbell’s heaviness, and my tactile experience of the balance of this heaviness on its two ends—seems centered upon a zero point that coincides with my entire body, which itself is also
explicitly represented tactiley. And all these tactile intentional aspects of experience are experienced tactiley as *unified*, as all being aspects of current total tactile what-it-is-like-ness for me. The experiential zero point of this tactile dimension of the “unity of apperception” seems, once again, to coincide with my entire body—a body that also can be *explicitly* represented tactiley, via that ways I now tactiley perceive various of its parts.

Another dimension of subtlety is the way that zero-point aspects of sensory-perceptual experience arise multi-modally, often in combination with multi-modal aspects of explicit sensory-perceptual representation. For instance, the up/down axis of one’s visual-presentational experience seems heavily dependent upon certain aspects of tactile and kinesthetic sensory-perceptual experience: roughly, down is the direction of the surface to which one’s body at rest is attached, and toward which one’s body moves when unattached to any surface. This up/down axis affects the explicit representational content of visual experience, which represents objects in the ambient environment as located within an implicit reference frame that includes the directions of up and down. The up/down axis also is a zero-point aspect of visual-perceptual experience: the experiential self is directly present in visual experience as visually located not only in the spatio-temporally “absolute ‘Hier’” manner, but also in the (multi-modal) “up-vs-down” manner. (Husserl’s mathematical metaphor already strongly suggests this, of course: mathematical “up” in a three-dimensional Cartesian coordinate system is the positive-number direction on the z axis, and “down” is the negative-number direction on this axis.)

Much more could be said about such matters. But enough has been said for our purposes here. One main lesson is that sensory-perceptual experience is richly suffused with zero-point aspects—aspects that often are subtly multi-modal, and that often are subtly intertwined with explicit representation in sensory-perceptual experience. And, as emphasized in section 2.2, zero-point aspects of consciousness—including zero-point aspects of sensory-perceptual experience—accrue to the direct presence in consciousness of the experiencing self, a self-presence not mediated by experiential representations of the self. One experiences *being* located at a specific sensory-perceptual zero point, rather than experiencing representing oneself as being thus located.4

2.5. Voluntary-Action Experience

---

4 Some philosophers maintain that sensory-perceptual experience does not involve representations at all, but instead has intentionality consisting of direct experiential acquaintance, unmediated by representations, with various objects in one’s ambient environment and with instantiations by those objects of various properties and relations (e.g., Campbell 2002, Hellie 2007, Nudds 2009). In our opinion, however, the falsity of such “naïve realism” about sensory-perceptual experience should be virtually self-evident, among other reasons because advocates of the position are forced into wildly implausible “disjunctivism” about the intentionality of hallucinatory sensory-perceptual experience vis-à-vis the intentionality of phenomenologically indistinguishable sensory-perceptual experience that is veridical. The kernel of truth in so-called naïve realism is that objects in one’s ambient environment are experientially *presented* in sensory-perceptual experience, when such experience is not illusory or hallucinatory; and being thus presented is a form of sensory-perceptual *acquaintance* with those objects (and with their perceived properties and relations). But the mistake is to construe sensory-perceptual acquaintance as non-representational. Rather, for objects to be presented by sensory-perceptual experience, and to be presented as instantiating various properties and relations, is, roughly, a matter of (i) one’s current sensory-perceptual experience presentationally representing such putative objects, and presentationally representing such putative objects as instantiating those properties and relations, (ii) such objects’ being now actually present in one’s ambient environment, (iii) those objects’ instantiating close enough (if not perfectly) the properties that one’s current sensory-perceptual experience presentationally represents such putative objects as instantiating, and (iv) one’s current sensory-perceptual experience, as-of such objects instantiating these properties and relations, being caused by those very objects’ instantiating (close enough) those very properties and relations.
In this subsection we begin by describing some features of agentive phenomenology which, we submit, are readily ascertainable just on the basis of introspective attention to such phenomenology. We then urge that these features are best characterized as zero-point aspects of phenomenal subjectivity, rather than as involving explicit self-representations.

What is behaving like phenomenologically, in cases where you experience your own behavior as action? Suppose that you deliberately do something—say, holding up your right hand and closing your fingers into a fist. What can you ascertain about the phenomenology of this item of behavior, on the basis of introspective attention to this phenomenology? To begin with, there are of course the purely bodily-motion aspects of the phenomenology—the what-it’s-like of being visually and kinesthetically presented with your own right hand rising and its fingers moving into clenched position. But there is more to it than that, of course, because you are experiencing this bodily motion as your own action.

In order to help bring into focus this specifically actional phenomenological dimension of the experience, it will be helpful to approach it a negative/contrastive way, via some observations about what the experience is not like. For example, it is certainly not like this: first experiencing an occurrent wish for your right hand to rise and your fingers to move into clenched position, and then passively experiencing your hand and fingers moving in just that way. Such phenomenal character might be called the phenomenology of fortuitously appropriate bodily motion. It would be very strange indeed, and very alien.

Nor is the actional phenomenological character of the experience like this: first experiencing an occurrent wish for your right hand to rise and your fingers to move into clenched position, and then passively experiencing a causal process consisting of this wish’s causing your hand to rise and your fingers to move into clenched position. Such phenomenal character might be called the passive phenomenology of psychological state-causation of bodily motion. People often do passively experience causal processes as causal processes, of course: the experience of seeing the collision of a moving billiard ball with a motionless billiard ball is an experience as-of the collision causing the latter ball’s subsequent motion; the experience of observing the impact of the leading edge of an avalanche with a tree in its path is an experience as-of the impact causing the tree to become uprooted; and so on. Sometimes people even experience their own bodily motions as state-caused by their own mental states—e.g., when one feels oneself shuddering and experiences this shuddering as caused by of a state of fear. But it seems patently clear that one does not normally experience one’s own actions in that way—as passively noticed, or passively introspected, causal processes consisting in the causal generation of bodily motion by occurrent mental states. That too would be a strange and alienating sort of experience.

How, then, should one characterize the actional phenomenal dimension of the act of raising one’s hand and clenching one’s fingers, given that it is not the phenomenology of fortuitously appropriate bodily motion and it also is not the passive phenomenology of psychological state-causation of bodily motion? Well, it is the what-it’s-like of self as source of the motion. You experience your arm, hand, and fingers as being moved by you yourself—rather than experiencing their motion either as fortuitously moving just as you want them to move, or passively experiencing them as being caused by your own mental states. You experience the bodily motion as generated by yourself.

---

5 This first part of this subsection is adapted, with some modifications, from similar discussions in several other texts including Horgan, Tienson and Graham (2003) and in Horgan (forthcoming).

6 Phenomenologically speaking, exercises of one’s agency are experienced as temporally located, and in that respect are experienced as “event-ish.” Thus, the expression ‘state causation’ works better than ‘event causation’ as a way of expressing the way behaviors are not presented to oneself in agentive experience. (States can be short-lived, and often when they do they also fall naturally under the rubric ‘event.’) Although agentive experience is indeed “event-ish” in the sense that one experiences oneself as undertaking to perform actions at specific moments in time, one’s behavior is not experienced as caused by states of oneself.

7 For discussion of a range of psychopathological disorders involving similar sorts of dissociative experience, see Stephens and Graham (2000).
The phenomenology of doing typically includes another aspect, which we will call \textit{core optionality}. (More presently on the reason for the modifier ‘core’.) Normally when you do something, you experience yourself \textit{as freely} performing the action, in the sense that it is \textit{up to you} whether or not to perform it. You experience yourself not only as generating the action, and not only as generating it purposively, but also as generating it in such a manner that you \textit{could have done otherwise}. This palpable phenomenology of optionality has not gone unrecognized in the philosophical literature on freedom and determinism, although often in that literature it does not receive as much attention as it deserves. (Sometimes the most explicit attention is given to effort of will, although it takes only a moment’s introspection to realize that the phenomenology of effortfully exerting one’s will is really only one, quite special, case of the much more pervasive phenomenology of optionality.)

The core-optionality aspect of agentive phenomenology is intimately bound up with the aspect of self-as-source, in such a way that the former is an essential component of normal agentive self-source experience. In experiencing one’s behavior as emanating from oneself as its source, one experiences oneself as being able to refrain from so behaving—or at any rate, as being able to refrain from willfully producing such behavior. This is so even when one acts under extreme coercion or duress—e.g., handing over one’s wallet or purse to a thief who is pointing a gun in one’s face. It also is so even when one acts with an extreme phenomenological “imperativeness”—e.g., a mother’s unhesistatingly leaping into the river to save her drowning child, Luther’s acting out a sense of moral requirement (as expressed by his declaring “Here I stand, I can do no other”), or the compulsive hand-washer’s act of washing hands for the third time in ten minutes. The \textit{core} phenomenology of optionality that is essential to ordinary agentive experience remains present in all such cases, even though there are further, superimposed, phenomenological aspects (duress, moral-obligation experience, intensely strong irrational desires, or the like) whose presence can render appropriate, in context, a judgment that the agent “could not have done otherwise,” or “had no other option,” or “did not act freely.” (Such phenomenological aspects fall under a rubric we will call \textit{superimposed non-optionality}.) Because the phenomenology of core optionality remains present even when the phenomenology of superimposed non-optionality is also present, it can be contextually appropriate to use ‘could’ and ‘option’ and ‘free’ in a way that reflects this fact (rather than in a way that reflects the presence of one or another kind of superimposed non-optionality phenomenology). For instance, one might say this: “I could have refrained from giving the thief my wallet, and thus I gave it to him freely and with the option of refraining—even though refraining would have been quite stupidly irrational.”

Agentive phenomenology is more closely akin to perceptual/kinesthetic experience than it is to discursive thought. (Many higher non-human animals, we take it, have some agentive phenomenology, even if they engage in little or no discursive thought.) Of course, we humans also wield \textit{concepts} like agency, voluntariness, and the like (whereas it is questionable whether non-human animals do); but thoughts employing these concepts are not to be conflated with agentive phenomenology itself.

Our remarks thus far in this subsection are offered as the products of attentive phenomenological introspection. Pay attention to your own agentive phenomenology, we submit—perhaps while also contemplating intelligible-seeming contrast-cases, like the (imagined) phenomenology of fortuitously appropriate bodily motion and the (imagined) passive phenomenology of psychological state-causation of bodily motion—and you should find introspectively that your agentive experience has the features we

---

\footnote{This is not to deny, of course, that there is indeed a distinctive phenomenology of effort of will that \textit{sometimes} is present in the phenomenology of doing. The point is just that this aspect is not always present.}
have been describing. The following question now arises. Are these features zero-point aspects of phenomenal subjectivity, or do they instead require, and deploy, an explicit self-representation?

An apt way to address this question is to recur to a consideration already emphasized above, viz., the fact that the experiential self is present in consciousness more intimately—more directly—via phenomenal subjectivity than it ever is—or ever can be—by being explicitly represented. With that in mind, ask yourself introspectively which of the following two descriptions better fits the fundamental character of your own agentive experience:

(A) My agentive experience is a matter of me myself being explicitly represented in consciousness, and moreover is a matter of me myself being explicitly represented as being the source of my action and as possessing core optionality in producing it.

(B) My agentive experience, rather than being a belief-like matter of explicitly representing myself, to myself, as the core-optionality possessing source of my action—and rather than being a matter of undergoing (via a presentational representation) an explicit, perceptual or quasi-perceptual, presentation of myself, to myself, as the core-optionality possessing source of my action—instead is the direct experience of being the source of my action, and of exercising core optionality.

Description (B), we submit, is the better one. The intertwined experiential aspects of self-sourcehood and core optionality are, most fundamentally, aspects of the directly experienced presence of the self in consciousness—aspects of the for-me-ness of agentive experience. Thus, they are zero-point aspects. Although we humans also can undergo conscious states—e.g., beliefs or propositional entertainings—that deploy explicit self-representations and that explicitly attribute self-sourcehood and core optionality to this self, this is decidedly an “add-on,” something over and above agentive experience itself. (And surely creatures like dogs cannot undergo mental states of that kind, even though it is plausible that dogs too undergo, to some extent at least, agentive phenomenology with its aspects of self-sourcehood and core optionality.)

What about presentational representations, of the kind deployed by sensory-perceptual experience? Well, what introspection seems to reveal is that the sensory-presentational aspects of agentive experience are all, and only, those that would also be possessed by experiences with the phenomenology of fortuitously appropriate bodily motion—viz., being perceptually presented (e.g., visually and kinesthetically) with various aspects of one’s body’s motion. Self-sourcehood and core optionality are not presented in experience the way one’s bodily motions are presented; rather, they are present more directly and intimately than that. They are zero-point aspects of phenomenal subjectivity.

2.6. Prospection and Recollection

Ordinary conscious experience typically is suffused with prospective elements. As one moves purposefully through one’s ambient environment, for example, one constantly anticipates one’s upcoming sensory-perceptual experience in the immediate and short-term future, and one constantly modulates the specifics of one’s behavior in light of various prospective elements. Walking with proper balance, for instance, depends in part upon anticipated upcoming impact-cum-support experience vis-à-vis the leg and

---

9 Although it is should be obvious to introspection that agentive experience has these features, this does not mean that it is obvious to introspection what constitute the satisfaction conditions of self-sourcehood and core optionality. For arguments in support of the claim that introspection alone cannot reliably ascertain these satisfaction conditions, see Horgan and Timmons (2011) and Horgan (2012, forthcoming).

10 Actually, kinesthetic experience sometimes presents parts of one’s ambient environment through what might be called “extended touching”—e.g., feeling the leaves one is currently raking via how the rake feels, feeling the position of one’s front wheels as one guides one’s automobile through a turn, etc. But this does not affect the point we are making in the text.
foot that one is now extending forward. Purposefully pushing one’s grocery cart down specific aisles in the grocery store (while purposefully bypassing others), for instance, depends in part upon anticipated sensory-perceptual experiences of being visually-perceptually presented with specific items that one intends to purchase, situated on a grocery shelf immediately in front of oneself and within reach. Such examples are legion.

Similar remarks apply to recollective elements of experience. Walking with proper balance depends in part upon recently experienced, recollected, impact-cum-support experience vis-à-vis the leg-cum-foot that just recently was supporting one’s body and is now moving forward. Purposefully pushing one’s grocery cart down specific aisles depends in part upon recently experienced, now recollected, experiences of seeing and removing certain specific items from certain shelves. Again, such examples are legion.

The experiencing self is experientially present, in such prospective (or recollective) experiences, in a direct and immediate way: one is envisioning, or remembering, a certain what-it-is-likeness for me. Indeed, the experiencing self is directly and immediately present twice over, so to speak. The phenomenal character for me now, of my current prospective (or recollective) experience involves an envisioned (or recalled) phenomenal character for me then. Introspective attention to such prospective/recollective elements of experience thus reveals that these elements have temporally layered, or temporally superimposed, zero-point aspects: there is something it is like, for me now, to envision (or recollect) a something-it-will-be-like (or a something-it-was-like) for me then. And these are indeed zero-point aspects, rather than involving explicit conscious representation of oneself to oneself. For, to repeat, the experiencing self is directly and immediately present in such experiences, twice over.

2.7. Desire

One’s occurrent conscious desires frequently pertain to oneself: what one desires is some state of affairs involving oneself—say (upon having just finished drinking the beer one has been holding), to be in the possession of another glass of beer, and to drink that beer. The following question now arises. In typical cases of undergoing such a self-directed, occurrent, desire (e.g., the desire to drink another beer), does one consciously and explicitly represent oneself, to oneself, as being in that desired situation, or does one rather experiencing a desire-attitude vis-a-vis an envisioned state of affairs with a prospective-zero-point aspect?

An apt way to address this question is to ask whether the prospective involvement of oneself, in the envisioned situation that one desires to come about, is experientially direct and immediate. Introspective attention to such desire-experiences reveals that the answer is Yes. What one envisions, and envisions in the desiring-for way, is to be drinking another beer. One is envisioning occupying the sensory-perceptual zero point of a beer-drinking experience, and one is envisioning this (desire-ingly) while yet also occupying the zero point of a current sensory-perceptual experience that presents to oneself an empty glass in one’s hand. Thus, occurrent conscious desires of this kind have the twice-over zero-point aspect we described in section 2.4. The experiential self is involved more immediately and directly in such experiences than it would be, or could be, through mediation via a conscious representation of oneself drinking beer.

This is not to deny that people can, and sometimes do, undergo occurrent, conscious, self-involving desires that do involve conscious self-representation. This can happen, for instance, when one’s desire rests upon a distanced perspective upon oneself, such as a third-person perspective. Perhaps

---

11 Extremely short-term cases of such prospection and recollection perhaps coincide with what Husserl called the retentive and protentive aspects of the “specious present.” But presumably other cases do not, such as envisioning or recalling removing items from shelves that are located several aisles away from one’s current location in the grocery store.
one actually does not desire to drink another beer (one feels beer-satiated); but one does desire to appear hip to others in the room, and one believes that drinking another beer will have this effect. One thereby forms a desire, vis-à-vis a state of affairs consisting of one’s drinking another beer, in which this state of affairs is envisioned from a third-person perspective: one consciously, desire-ingly, represents oneself, to oneself, as seen by others, drinking beer and thereby looking hip to them. As one might express in words one’s conative situation with respect to drinking another beer, “I don’t desire to drink another beer, but I do desire that I drink another beer.”

Ordinary terminology for describing ordinary self-involving desires is suggestive and illuminating here, as is the locutionary awkwardness of describing such desires the way philosophers typically do when treating them as so-called “propositional” attitudes. The ordinary mode of speech is to use an infinitival construction, without any first-person pronoun—as in “I want to drink another beer.” Philosophers are often tempted to think that a more philosophically revealing way to express one’s desire would be propositionally, using an explicit first-person indexical term—as in the awkward, stilted-sounding, “I want that I drink another beer.” Now, although one must certainly be cautious in seeking to extract philosophical morals from ordinary-language grammar, in the present case the facts of English grammar do seem to provide further support for what we are claiming (and we suspect that many other languages are similar in this respect). The ordinary locution seems especially apt as an expression of self-involving desires that have the twice-over zero-point aspect and are not explicitly self-representational. The awkward, stilted-sounding, locution, does seem at least somewhat apt, however, as a device for expressing an explicitly self-representing desire involving a third-person perspective upon oneself.

2.8. Reasons

Intention formations and actions normally are highly purposive, and normally exhibit practical rationality vis-à-vis the purposes that lie behind them. The paradigmatic structure of reasons-based intention formation and action is the familiar desire/belief structure: one has a certain desire (often a self-involving one) that is not overridden by conflicting desires, one has a belief to the effect that a specific available potential action would (probably) lead to the fulfillment of the desire, and straightaway one either performs such an action or forms an intention to do so later. Typically, the pertinent desire and the pertinent belief are each occurring and conscious, although in some cases (e.g., performing highly routine actions like unlocking one’s office door) they are not at the center of one’s attention.

We will henceforth use the locution ‘voluntary act’ to cover both actions (including mental actions, such as mental addition) and intention formations. We take it that in cases where one performs a specific voluntary act because of an occurring, conscious desire/belief pair that together constitute a reason for such an act, the belief and desire jointly state-cause the act. On the other hand, as we emphasized in section 2.5, one normally does not experience one’s act as being thus state-caused. Rather, and again as we stressed in section 2.5, one’s voluntary-act experience has the zero-point aspect of self-sourcehood. But in addition, one’s experience also has two further pertinent elements: first, one experiences such an act, considered prospectively, as rationally favored by the pertinent desire and belief; and second, in voluntarily performing the act, one experiences doing so on the basis of this favoring relation. And, once again, all this is not experienced, passively, as a state-causal process involving occurrence conscious mental states state-causing bodily motions or intention formations; rather, it is

12 Our thanks to Mark Timmons for a conversation (over beer!) that prompted the present paragraph.
13 But note well: not experiencing one’s voluntary act as state caused is different from, and weaker than, experiencing one’s act as not state-caused. It is a vexed question whether or not the latter feature is an aspect of ordinary agentic experience, and it is also a vexed issue whether or not direct introspection can reliably ascertain the answer to this vexed question.
experienced actively and agentively. One experiences oneself as exerting voluntary agency, and as doing so rationally—i.e., as doing so on the basis of reasons that one experiences as favoring one’s act. (Often, especially in performing routine acts like brushing one’s teeth, the phenomenological features we are describing are outside the focus of attention; but they are present nonetheless.)

Wilfrid Sellars described in striking fashion a distinction of the kind we are now highlighting: the distinction between “the space of causes” and “the space of reasons.” Phenomenologically, voluntary-agency experience involves occupying a specific locus in the space of reasons. That locus is determined by various aspects of one’s own current experiential situation—in particular, one’s present occurring desires, and one’s present occurring beliefs concerning potential desire-implementing acts one might perform.

The following question now arises. In ordinary cases where one exercises one’s voluntary agency, does one consciously and explicitly represent oneself, to oneself, as occupying a specific locus in the space of reasons, or is it rather the case that being thus located is a zero-point aspect of consciousness? As with previous such questions addressed above, an apt way to address this question is to ask whether the reasons-appreciating, reasons-as-basis, elements of one’s voluntary-agency are experienced in an immediate and direct way in one’s voluntary-agency experience. The answer delivered by attentive phenomenological introspection is Yes. One’s experience is as-of occupying a specific locus in the space of practical reasons: (i) of immediately and directly having those reasons, (ii) of immediately and directly appreciating the favoring relation between those reasons and one’s voluntary act, and (iii) immediately and directly of performing this act on the basis of those reasons.

This is not to deny that people can, and sometimes do, act voluntarily on the basis of reasons constituted by occurring conscious states some or all of which involve conscious, explicit, self-representation. Once again, however, voluntary acts of that kind typically rest upon a distanced perspective on oneself, such as a third-person perspective. Imagine someone, for instance, who has absolutely no desire to exercise at the local gym several days per week, but who does desire to lose weight. Although he immediately and directly believes that gym exercise is the most effective way to lose weight, and he immediately and directly desires to lose weight, his loathing of exercise is so intense that no immediate, direct, desire to exercise arises within him. Nonetheless, he is able to take a third-person perspective on himself, and to summon up a desire with respect to a state of affairs as envisioned in the explicitly self-representational way—a state of affairs envisioned via consciously deploying a self-representation, aptly expressible as “I go to the gym and exercise.” Whenever it is gym time, he summons up this self-representing desire, perhaps by imagining seeing in the mirror a slimmer, trimmer, reflection than he actually still sees every day in the mirror. So, even though he has no desire whatever to go to the gym, he goes anyway because he is motivated by the self-representing desire that I go to the gym.

Thus far in this subsection we have been discussing the phenomenology of voluntary agency, and of the associated reasons-involving aspects of practical rationality. We have deployed the expression ‘voluntary act’ broadly, to encompass voluntary intention- formations as well as voluntary actions. But the space of reasons figures not only in the phenomenology of experiences involving practical rationality, but also in the phenomenology of experiences involving theoretical rationality. A central phenomenon in this domain is belief fixation—i.e., formation of, or maintenance of, beliefs. Most philosophers, ourselves included, maintain that belief fixation is not under voluntary control; one major reason to think so is that this attentive introspection reveals it to be so. Nevertheless, there is the following, phenomenologically

---

14 One’s own current experiential situation might also have a phenomenal character that involves implicit appreciation of certain background desires and/or background beliefs—desires and beliefs that are not presently occurring. For discussion of such “chromatic illumination” of conscious experience by such “morphological content,” see Horgan and Potrč (2010).
palpable, respect in which episodes of belief formation deserve to be classified as mental acts—albeit involuntary ones. One’s experiences are as-of being a rationally responsible epistemic agent—an agent capable of appraising and evaluating one’s available evidence, capable of appreciating the extent to which that evidence favors or disfavors various potential beliefs, and capable of forming beliefs on the basis of sufficiently strong evidential support. Such experiences are not passive, in the manner of sensory-perceptual experiences—even though they are not voluntary either. Rather, rational evidence-assessment and rational belief-formation are experienced as mental acts one consciously performs—acts consisting of exercising one’s rationality.

So the following question now arises, similar to others raised above. In ordinary cases where one exercises one’s epistemic agency, does one consciously and explicitly represent oneself, to oneself, as occupying a specific locus in the space of reasons, or is it rather the case that being thus located is a zero-point aspect of consciousness? Once again, an apt way to address this question is to ask whether the reasons-appreciating, reasons-as-basis, elements of one’s (involuntary) exercises of epistemic agency are experienced in an immediate and direct way in one’s evidence-appraising and belief-forming experience. The answer delivered by attentive phenomenological introspection is Yes. One’s experience is as-of occupying a specific locus in the space of epistemic reasons: (i) of immediately and directly having those reasons, (ii) of immediately and directly appreciating the favoring relation between those reasons and one’s potential belief, and (iii) immediately and directly (albeit involuntarily) forming this belief on the basis of those reasons. Thus, the distinctive phenomenology of reasons-responsive exercises of epistemic agency constitutes yet another zero-point aspect of phenomenal subjectivity.

This is not to deny that people can, and sometimes do, rationally form (or maintain) beliefs only in a way that is non-immediate and non-direct, and that deploys an explicit conscious self-representation. Suppose, for example, that one remembers appreciating that the specific considerations $C_1, C_2, \ldots, C_n$ confer strong evidential support on proposition $P$, and one remembers forming the belief that $P$ on this basis, and yet at the present moment one is unable to generate an occurrent appreciation of why it is that $C_1, C_2, \ldots, C_n$ confer strong evidential support on $P$. (Perhaps this inability is the result of having recently consumed too much beer.) One can rightly say to oneself, “I believe that $C_1, C_2, \ldots, C_n$ confer strong evidential support on $P$,” and one can legitimately persist, on this basis, in believing $P$. But this explicitly self-representational way of believing $P$ on the basis of $C_1, C_2, \ldots, C_n$ is a decidedly indirect way of being self-involved with respect to the reasons for $P$. Ordinary, full-fledged, evidence appraisal and evidence appreciation are are zero-point aspects of experience, with the pertinent zero point being one’s own present locus within the space of epistemic reasons.

2.9. The Phenomenal Character of Explicit Self-Representation

A general lesson that emerges from the preceding discussion is that explicit self-representations are apt to figure in conscious experience when one regards oneself from some kind of distanced perspective, such as a third-person perspective. Of course, the aspect of phenomenal subjectivity will still be present, since it is integral to all conscious experience: a conscious episode that includes explicit self-representation will still possess for-me-ness. Also, it seems entirely possible that some explicitly self-representational experiences will include yet another phenomenally subjective aspect: one’s experience might represent some self-involving episode (with the self being experientially present in the immediate and direct way, non-representationally), while also incorporating a reflective, third-person, perspective toward oneself in that very episode.

Consider, for instance, experiences of regret. Phenomenological introspection suggests that often such an experience includes both elements lately mentioned. On one hand, one recollects being at the agentive zero-point and the sensory-perceptual zero-point of some specific past action. On the other hand,
one also is taking a critical, distanced, perspective upon oneself—which involves explicitly representing oneself, to oneself, as having behaved badly in so acting.

It is to be expected, then, that instances of explicit self-representation will tend to arise when the cognitive task at hand calls for some kind of distanced perspective upon oneself. And it is also to be expected that sometimes such distanced, self-representing, consciousness regarding some recollected (or prospective) self-involving episode will be experientially superimposed upon the direct, non-self-representing, recollection (or prospection) of that very same episode.

Even if this is right, however, explicit self-representation also is apt to arise under other various other kinds of circumstance too—in particular, when the cognitive task at hand can be more easily, more resource-efficiently, accomplished by deploying I*/me* representations. Indeed, and notwithstanding our remarks in section 2.6 above, explicit self-representations often can enhance and streamline the cognitive-architectural operation of both episodic memory and planning—a theme we pursue below.


3.1. Preliminaries

As we said in section 2.1, the inherent phenomenal character of a particular mental state or process is the experiential categorical basis of the pertinent functional-role aspects of that state/process—aspects that are appropriate to its intentional content. In the case of self-involving conscious states/processes—both those involving only zero-point aspects and those that also deploying self-representations—content-appropriateness of functional role normally will be relativistic: the functional role played by the given state/process, in the ongoing cognition-cum-behavior of the person experiencing that state/process, will be content-appropriate given that the state/process is being experienced by that same person. Such experience-relativity of content-appropriateness will obtain not only for conscious states/processes in which the self is explicitly represented, but also for ones in which the self is present only immediately and directly via zero-point aspects.

Conscious states/processes are implemented—i.e., realized, or subserved—by states/processes that are characterizable in sub-phenomenological ways. Implementation, in turn, is a hierarchical matter: it involves various successive theoretical-descriptive levels, bottoming out at the level of subatomic physics. One implementational level, high up in the hierarchy and just below the level of conscious experience itself, is the level of what is sometimes called cognitive architecture. Much theorizing in cognitive science occurs at this level. Here the goal is to describe in a fairly abstract way nature’s “engineering design” for cognitive agents, and for humans in particular. Theoretical hypotheses are put forward about abstract functional organization, involving notions like “central processing unit” and “belief box.” Theoretical hypotheses are also put forward about the internal structure of the cognitive-architectural-level states/processes that implement conscious states/processes—for instance, the hypothesis that they all have digital, language-like, structure and belong to a “language of thought”, or the hypothesis that they all have analog, picture-like structure, or the hypothesis that some consciousness-implementing states/processes are language-like and others are picture-like.

Although the inherent phenomenal character of a given conscious state/process is indeed the experiential categorical basis of the pertinent functional role of that state/process, this functional role also will have successive implementing categorical bases as well—at the level of cognitive architecture, at the neural level, and ultimately at the subatomic level. Our concern in the remainder of this paper is with cognitive-architectural implementation of various kinds of conscious states/processes. In particular, we are concerned with (i) when, and why, such implementation involves representations with an explicit
I*/me* constituent, and (ii) when, and why, the implementing representations instead have zero-point structure.

The deliverances of attentive phenomenological introspection, we take it, can be rightly treated as providing strong—albeit defeasible—evidence about this matter. Conscious states in which the experiencing self is present immediately and directly, and with various zero-point aspects, can be expected to be cognitive-architecturally implemented via representations with zero-point structure. And conscious states whose content is overtly self-representational can be expected to be cognitive-architecturally implemented via representations that include an explicit I*/me* constituent. The reason why the phenomenological evidence is defeasible is this: in principle, conscious states/processes might get cognitive-architecturally implemented via representations some of whose structure is not reflected within consciousness itself. For instance, in principle the zero-point aspects of conscious experience could all be implemented by representations that deploy an I*/me* constituent sub-consciously, even though these zero-point aspects do not involve conscious self-representation. But the theoretical burden of proof rests upon those who would make such claims; meanwhile, phenomenology can and should inform theorizing about the nature of the representations that cognitive-architecturally implement conscious states/processes. In addition, further evidence can be garnered from empirical data, and from pertinent extant theory, within cognitive science itself; some of this data and theory will be invoked below.

What might zero-point representational structure be like? Various possibilities suggest themselves. Picture-like analog representations, for instance, have built-in zero-point structure: they represent a scene from a point of view—a point of view that does not get represented itself. Language-like representations can deploy representational categories that presuppose an implicit frame of reference that is centered on oneself, and can do so without also deploying any I*/me* representational constituent—categories like “above and to the right.” Also, some language-like representations might have a structure other than that of a declarative sentence, a structure that nonetheless is suitable for their self-involving functional role—e.g., representations that have the grammatical structure of commands. For instance, representations that have a command structure, and lack any language-of-thought analog of a grammatical subject, could contribute to intention-formation by getting inserted into a cognitive-architectural “intention box.” So there is certainly no in-principle obstacle to the possibility that certain representations can play self-involving roles in a creature’s cognitive economy without containing an I*/me* representational constituent.

3.2. Cognition without Me*

3.2.1. Judgment and Decision Making

---

15 Questions about how various kinds of phenomenally conscious states/processes get implemented, including questions about how they get cognitive-architecturally implemented, are themselves tractable, regardless of what one might think about the tractability or intractability of the so-called “explanatory gap” (Levine 1983) or “hard problem” of phenomenal consciousness (Chalmers 1996)—roughly, the question of why it should be that so-and-so physical or functional state/process should be a supervenience base for such-and-such phenomenal character, rather than instead either (i) being a supervenience base for some other phenomenal character, or (ii) occurring without any accompanying phenomenal character at all.

16 It is possible in principle, and may well be possible in practice, for there to be critters—say, certain kinds of robots—that (i) are correctly describable cognitive-architecturally as deploying both zero-point representations and I*/me* representations, but (ii) are “zombies” with no phenomenology at all (and hence with no zero-point aspects of phenomenology, and with no explicitly self-representational, essentially indexical, aspects of phenomenology). Much of what we say hereafter might well be applicable, mutatis mutandis, to such zombie critters, although none of our phenomenological discussion in section 1 would apply to them.
In his textbook, Rey draws on the ideas of Perry and Castañeda to argue explicitly for a mental item in the language of thought corresponding to the natural language term “I”:

Just as in English the speaker is supposed to use "I" only to refer to his or herself, in LOT the "system" uses a certain term to refer only to the receiver of present inputs, the instigator of outputs, and the subject of intervening mental states.

Rey goes on to elaborate:

Suppose the term were "i," and that the system uses "i" to record automatically that it had certain perceptions and judgments and preferences; and that its behavior is crucially determined by just those attitudes that do have this "i" as their subject: i.e. its actions are standardly caused by beliefs and preferences that are designated as belonging to "i." Merely my comp-judging the sentence "GR ought to get moving right now" won't be enough to get me moving; I've got to comp-judge "i had better get moving right now." (Rey 1997, p. 291, emphasis added)

Relatedly, Perry writes, “what is special about self-notions is that they are the normal repository of normally self-informative ways of perceiving, and the normal motivator of normally self-dependent ways of acting” (Perry 2002, 202).

These suggestions about the role of I* in action seem initially tempting. To be sure, Perry cases show how representing SN ought to move right now might not suffice to get SN moving where I* ought to move would suffice; similarly for TH. But it is a mistake to infer from this fact that the action of moving requires tokening the I* symbol, as in I* ought to move right now. It’s perfectly coherent to have a decision making architecture that effectively presupposes that perceptions are one’s own and actions are one’s own without having to explicitly represent I*.

Just to take one very prominent species of AI models of decision making, consider production systems. Production systems are composed of conditional statements, where the consequent is an action. If one of the antecedents is represented as satisfied (in a working memory system), the program proceeds to carry out the act specified in the consequent. One simple example is provided by thermostats (e.g., Newell 1990). A thermostat might have the following production system:

\[
\text{If temperature} > 78 \text{ degrees and AC is off, turn AC on}
\]
\[
\text{If temperature} < 77 \text{ degrees, and AC is on, turn AC off}
\]

As in this example, typically, production systems do not including anything like the I* symbol. There is a hard-coded connection between the action indicated in the consequent and the requisite behavior. The thermostat doesn’t need to specify “turn on my* AC”. There is only one AC that is being controlled by the thermostat.

Let’s consider now a perfectly reasonable (if very simple) production system for a familiar childhood winter activity – catching snowflakes on your tongue:

\[
\text{If there are snowflakes falling from directly above, then tilt head upwards, open mouth, stick out tongue.}
\]

Just as with the thermostat, this production rule has no I* symbol. But, one might ask, doesn’t the system have to represent that snowflakes are falling above me*; that it’s my* head, my* mouth, my* tongue? No. Just as the connection between the particular thermostat and the particular AC is hardwired into the system, so too the connection between the particular instance of the production rule and the particular head, mouth, etc. are hard wired into the system. In other words, the production rule only sends instructions to my head (it doesn’t connect to other heads), so the rule doesn’t need to include a further specification that it is my* head.

Unlike the thermostat, people can, of course, have a representation like lift my* head. This is an important fact about us. One way in which it’s important is that it allows us to convert ambiguous environmental information into self-relevant representations. All of Perry’s examples work like this. I see a trail of sugar, and I seek someone under the description “guy spilling sugar” (Perry 1979). Or I see a
frumpy guy in the mirror and I wonder who he is (Perry 2002). Or I am told that some individual is about to be attacked by a bear (Perry 1977). In each of these cases, in order to get the appropriate action, I have to realize that I am the guy. But in normal decision making, there isn’t any question who the guy is.

While Perry’s examples illuminate a key element in the human representational system, it’s easy to exaggerate the role the I* plays in everyday cognition. We suspect that for most human problem solving and decision making, people do not token the I*.

This is lent some support by concurrent protocol analysis, in which people report what is running through their mind as they are solving a problem. Although philosophers of psychology often emphasize ways in which introspection fails, a large tradition of unsung work shows that people are quite accurate at reporting how they solve problems using the “think aloud” technique. For instance, when people are asked to solve a simple arithmetic problem like 19*4 and “think aloud” while they do it, people are excellent at reporting the process. The accuracy of self-report is confirmed by independent measures (see Ericsson & Simon 1993). For example, those who report converting 19*4 to [(20*4)–4] finish faster than those who report multiplying 4 by 9, carrying the 3, then multiplying 4 by 1 and adding the 3 to that. Similarly, the accuracy of self-report is confirmed by the relation between the participant’s solution and the reported process. This is especially clear in cases where participants make mistakes. For instance, if a participant says that 19*4 is 56, we can see how they went wrong if they report an arithmetic error like 20*4=60. Thus, protocol analysis plausibly gives an accurate window into at least some critical aspects of cognitive phenomenology. When we turn to the issue at hand, the representation of the self in decision making, we find that when people think aloud during judgment and decision tasks, they rarely mention the self (except incidentally – “I guess”, “I would say”) For instance, in solving problems under uncertainty, here are some representative reports from subjects:

“He is most likely a nurse just because being well spoken and interested in politics and having lots of time has nothing to do with being a doctor, and he can be a very good nurse, and there are more nurses than there are doctors.”

“He is a doctor because he is a male and not a lot of nurses are male, and he is well off and invests a lot of time in his career.”

“I would say Paul is a doctor because he invests a lot of time in his career and that probably takes more time than being a nurse” (De Neys & Glumicic 2008, 1288).

Or, for a mathematical example, here is the protocol from a participant asked to multiply 36 by 24:

“ok 36 times 24, um, 4 times 6 is 24, 4, carry the 2, 4 times 3 is 12, 14, 144, 0, 2 times 6 is 12, 2,carry the 1, 2 times 3 is 6, 7, 720, 720, 144 plus 720, so it would be 4, 6, 864” (Ericsson & Moxley 2011, p. 94)

Here is a representative example from participants doing a logic problem:

“A says B is a knave, that’s either true or false. Keeping that in mind, B says that A and C are of the same type. So if A is telling the truth, C is also of A’s type, which is truth-telling - knights - A and C are both knights if B is telling the truth. If B is telling the truth and A is telling the truth, well, something, neither, not both of them can be right, because either A is correct about B’s being a knave, or . . . wait, this is getting confusing” (Rips 1989, p. 89)

Here is an example from a decision concerning shopping:

Similarly, for most perceptions people have, we suspect that the I* isn’t tokened. Perception delivers representations like “there is a shoe next to the bookshelf” free of any explicit I* representation.

The reaction time studies typically use “retrospective” techniques so that the result can’t be attributed to the length of time it takes to utter the process. So, immediately after solving the problem, participants are asked to describe what went through their minds as they were solving the problem. People who report having gone through 7 steps reliably took longer to solve the problem than those who report having gone through 4 steps.
“cucumbers are still too expensive / Bacon costs too much to get at all / I get the, the little tenderized ones [steaks], what are they called? These are the ones if they get below a dollar…” (Svenson 1979)

And here is an example from that beloved technique in decision theory, gambling decisions:

“All PL [probability of losing] is … is lower on A / The SL [sum to be lost] is higher on A than on B / such a choice / Since B has a much higher PW [probability of winning]/ I’ll take that.” (Svenson 1979).

Although “I” does appear in some of these concurrent reports, it doesn’t seem to be integral to the problem solving process in the examples (though it will be in other cases, as we’ll see in section 3.3). It’s not that people are generally averse to mentioning themselves when engaged in reporting their thoughts. On the contrary, when people are asked to explain their thought process immediately after it occurs (using a “retrospective” rather than “concurrent” think aloud method), people typically do include explicit self reference:

“My first thought was I saw the 5600 dollars and I saw the 80 dollars but the 5600dollars — there is still a chance for me to gain, it was really small (probability low), but you never know. That’s why I chose B because 5600 dollars is a lot of money (value high)”

“My first thought was with 20% chance of owing 900 dollars, that gives me 80% chance (recode probability) to not owe 900 dollars and I have pretty good chances (probability high)” (Cokely & Kelley 2009).

Thus, for concurrent reports of thought processes, it is highly typical for subjects not to mention the self. Earlier we pointed out that sensible and efficient cognitive architectures don’t need an I* representation for basic thought and action. These points fit together, of course. An efficient architecture for problem solving doesn’t require the I*, and people don’t mention the I when solving simple problems. We suspect that this is because the I* concept is simply not represented in these exercises. It is possible, of course, that people really are explicitly the I* concept even in these reasoning exercises. But as far as we can tell, there is no reason to think this is the case. Rather, we maintain, the I* concept doesn’t appear in the simple thought and actional processes that are ubiquitous in daily life.

3.2.2. Phenomenology of Agency

We turn now to the distinctive phenomenology of agency. One of the more interesting developments in explaining the phenomenology of agency comes from work on the comparator model (e.g. Frith 2012, Gallagher 2000). The basic idea is that in a normal case of acting, like raising one’s hand in Q&A, when one forms the intention to raise one’s hand, this sends two signals forward. One signal (of course) goes to motor cortex to generate the appropriate muscular activity that eventuates in arm-raising; the other signal is called an “efference copy” and it predicts a certain outcome – one that conforms to arm-raising. This latter process is sometimes called a “forward” model because it’s prior to the actual movement. So, we have a motor signal that should lead us to raise our arm and a predictive signal that anticipates certain activity associated with arm-raising. A further mechanism, the comparator, evaluates whether the predicted activity actually occurs. Roughly, the comparator determines whether the arm has moved in the right way and at the right time given the prediction signal. If there is a match, then one will have the phenomenology of agency; otherwise one will not have that phenomenology.

The basic idea of the comparator model was developed earlier in animal models. Elephantfish are weakly electric fish that send out electrical discharges to gain information about their environment, including the location of other fish. But if they actually registered their own electric discharges, whenever the fish emitted a discharge, it would, as it were, think that it is right next to another fish. The fish need to discount their own electrical discharges to avoid all these false positives. The reigning model of how they

Note that subjects were instructed to begin their retrospective reports with “my first thought was”.
do this is a comparator model. When the command to emit a discharge goes out to the electric organ, an efference copy also goes out to the comparator. When there is a match between the predicted discharge and the registered electricity, the electricity is discounted from the inferences about the environment (Gandevia 1996).

The comparator account thus has impressive precedents in animal models. In the present case, the comparator account of the phenomenology of agency helps to explain why one can’t tickle oneself – the tickling phenomenology is defeated by the forward model which predicts the sensations. The account also explains why acting has a phenomenology different from being acted upon. If someone else pushes my hand up, the forward model isn’t generated and so I don’t have the comparator-generated phenomenology. The comparator account has also been invoked to explain clinical symptoms like anarchic hand in which people generate finely tuned motor behavior but lack the normal phenomenology of action. Relatively, the comparator account has been used to explain delusions of control and thought-insertion in schizophrenia (e.g., Wolpert & Ghahramani 2000).

It’s not clear how much of a reductionist one should be about all of this. But perhaps it’s worth noting that neurons in motor areas have axon collaterals that get sent to the cerebellum. (This means that in addition to the main projection of the axon, it has an additional projection that goes somewhere else. The tempting inference in this case is that when an instruction is issued by the neurons in motor cortex, the axon collaterals of those very neurons send an efference copy to the cerebellum.) And it’s known that patients with damaged cerebella exhibit behavior indicating reduced predictions of their own movements (Haggard & Whitford 2004, 52).

Thus the comparator model gives us an account of why it feels different to act vs. be acted on. But notice that nowhere in this discussion has there been need to make recourse to a self-representation. Researchers on electric fish do not posit an I* representation to explain how the fish discounts its own electrical activity. Similarly, in the comparator model of action, the entire process is run without invoking explicit self-representations. Of course, there is still a critical theoretical role for “self” in these theories. It is only because the motor instruction “raise hand” was issued by me that the forward model is generated. The same motor behavior forced on my by someone other than me would not produce the relevant phenomenology. But “I*” need not be tokened in order to generate that phenomenology (see also Prinz 2011).

3.3. Cognition that Needs Me*

3.3.1. Memory

Although we’ve argued that the bulk of human cognition can proceed perfectly well without representing the I*, we think that there are important exceptions. In recollecting a past event, we often need to represent who it happened to. I need to represent who the guy is that received the gift. That will determine whether I should express gratitude. Similarly, I need to represent who the guy is that was insulted. That will determine whether I should express hostility.

There are thus functional reasons for representing the self when we engage in recollecting episodes from our lives. It can be critical to determining the appropriate course of action. In addition, when we turn to descriptive psychology, we find that it’s common ground that in episodic recollection, the self is explicitly represented. The idea here goes back at least to William James, who wrote, “Memory requires more than the mere dating of a fact in the past. It must be dated in my past. In other words, I must think that I directly experienced its occurrence” (1890/1950, 650). This characterization of episodic memory continues to this day. In a recent piece on children’s episodic memory, Harlene Haynes and colleagues summarize the prevailing view that episodic recollection is “accompanied by conscious
awareness that the event happened to “me” or will happen to “me” that does not accompany retrieval of other kinds of memories” (Hayne et al. 2011, 344).

Indeed, there are cases of dissociations in which brain-injured patients lose the sense that their past experiences happened to them, and this is phenomenologically quite strange to the patient (e.g., Klein & Nichols 2012). One patient, RB, describes the phenomenon as follows:

I could clearly recall a scene of me at the beach in New London with my family as a child. But the feeling was that the scene was not my memory. As if I was looking at a photo of someone else's vacation.

He described another example, this time from his University days:

I can picture the scene perfectly clearly…studying with my friends in our study lounge. I can ‘relive’ it in the sense of re-running the experience of being there. But it has the feeling of imagining, [as if] re-running an experience that my parents described from their college days. It did not feel like it was something that really had been a part of my life. Intellectually I suppose I never doubted that it was a part of my life…. But that in itself did not help change the feeling of ownership.

RB’s lack of the felt sense that the past experience happened to him suggests that the normal course of episodic memory does indeed carry with it a distinct sense that the event happened to me.

There are two ways that the I* could get attached to episodic memories. One possibility is that the I* is actually encoded into each of the experience memories one has stored, e.g., in hippocampus (and other regions critical to stored memories). Another possibility that we think more likely is that the I* is automatically attached to an experience memory as it is retrieved from the bank of experience memories. On this approach, what happens with R.B. is that he has the experience memories stored normally, but his retrieval process is defective and fails to attach the I* symbol normally. Of course, the functional reasons for having an I* attached to memory that we suggested earlier, still apply. When I retrieve a memory from my store of experience memories, I need the representation of the experience to specify who the experience happened to. Moreover, this representation needs to be in a common language such that I can easily represent distinctions between the things that I did, the things that Mark did, and the things that Mark and I both did.

### 3.3.2. Planning

In addition to its role in episodic memory, we maintain that the I* symbol plays a key role in efficient planning. Perhaps the most influential account of planning in recent philosophy and artificial intelligence is Bratman’s “planning theory of intention” (e.g., Bratman 1987, 1999). Although it is not entirely explicit in Bratman’s theory, there is reason to think that the symbol I* is essential to efficient planning.

Bratman’s theory begins with the obvious fact that we are resource-limited creatures. We can’t always be considering all the options before us. At any point, there are indefinitely many possible courses of action, and the world constantly changes in ways that bring about new courses of action, new consequences for old actions and so forth. This makes ideal practical reasoning impossible for creatures

---

20 We acknowledge an apparent tension between this view and our remarks in section 2.6 about prospection and recollection as zero-point aspects of experience. Nonetheless, whereas we find it plausible that relatively short-term episodic-memory experiences normally are cognitive-architecturally implemented by zero-point representations, we also find it plausible that longer-term episodic-memory experiences instead need to be cognitive-architecturally implemented by I*-deploying representations. Roughly, the further in one’s past is the recollected episode, the greater is the need for the cognitive-architectural implementation of the episodic memory to include an I*-constituent—thereby facilitating the functional role of a genuine self-involving memory, as opposed (for instance) to a mere self-involving imaginative episode.

21 Some of the following material is adapted from Nichols (2000).
like us. Bratman argues that one way that practical reasoning is eased for us is that we often commit to an intention and then stop deliberating all things considered about what to do. Once we commit to an intention, we no longer consider all the possible alternative courses of action, we just take things from the committed intention, and this commitment constrains and structures subsequent planning and decision making. A key part of Bratman’s theory is that once you commit to a plan, you don’t even consider options that are incompatible with the intentions that you’ve committed to – incompatible options are “filtered” out (Bratman et al. 1991). For instance, if I have committed to an intention to go to Germany on June 15th, then I don’t even consider the option of hiking in the Sonoran desert on June 16th. A second key feature of the theory is that committing to an intention structures one’s subsequent deliberations. Of particular importance is that once one commits to an intention, this will structure future deliberations about how best to execute the plan. So, once I’m committed to going to Germany, I generate the subplan that I must get my passport renewed. On Bratman’s account, this kind of filtering and structuring is a much more efficient way to plan for creatures like us than to attempt a constant calculation of utility maximization.

Martha Pollack extends Bratman’s model by exposing another way the decision making of a resource-bounded agent can be made more efficient – by “overloading” one’s intentions (Pollack 1991, 1992). The idea is that if an agent has a goal, she can try search for whether that goal can be met in the course of executing some plan that has already been adopted. Suppose I realize that I need to get milk for breakfast tomorrow. I can then consider whether I already have a plan that will take me close to a grocery store. If I already planned to go to the hardware store to make a key, I can “overload” this prior plan to include a trip to the nearby grocery store.

What we want to suggest is that the I* symbol can play a critical role in these aspects of planning. If I plan to go to Germany on June 15th, then it’s an asset to filtering that I represent that I* will be unable to hike in the Sonoran desert on the 16th. And unless I represent that I* am going to Germany, I won’t know that I need to get my* passport renewed. Similarly, unless I represent that I* will be at the hardware store, it’s difficult to see how I can overload that plan for something else that I will need to do. So, while simple action might not require an I* representation, efficient planning often does.  

As with memories, there are different ways that the I* could get attached to the representations of our intentions. The I* might be encoded directly into each intention in the “intention box”. Another possibility is that the I* gets attached through the retrieval process. As with memory, we think this retrieval-based story is more plausible, partly because it strikes us as a more efficient cognitive architecture (and partly because we are sympathetic to the idea that the representations in the “intention box” are imperative in form rather than declarative). Again, the functional reasons for thinking the I* symbol is important to efficient planning still apply on this retrieval-based story. I need to represent which things I intend to do, which things my wife intends to do, and which things she and I intend to do together. Representing these expected facts enables me to exclude various further paths of decision making and helps me to determine which plan should be overloaded – mine, hers, or ours.

4. Conclusion

The distinction between simple action and planning resolves an erstwhile apparent tension between what we have been saying in this section about planning and our earlier discussion of prospection in section 2.6. Whereas the kind of prospection that is frequently involved in simple purposive action can plausibly be implemented by zero-point representations, the more complex one’s planning task becomes the greater are the cognitive-architectural advantages of implementation via I*-deploying representations.
In section 2, we argued that phenomenal subjectivity—the direct, non-representational presence of the experiential self in consciousness as the me for which one’s conscious experience has phenomenal character—is richly suffused with various, interpenetrating, zero-point aspects. These include the zero-point aspects of sensory-perceptual experience (in its several modalities), the zero-point aspects of voluntary-action experience (including the self-sourcehood aspect and the core-optionality aspect), the temporally layered zero-point aspects that figure in at least some instances of prospection and recollection, the zero-point aspect of ordinary self-involving desires, and the zero-point aspect (within the space of reasons) of both practical and theoretical rationality.

In section 3, we took up the question of how self-involving conscious experiences of various kinds are cognitive-architecturally implemented. We argued that problem solving often proceeds without an I* symbol. In terms of cognitive architecture, it’s easy to accommodate this for simple problems. Recall the example of production systems. The thermostat doesn’t need to specify my* AC because the connection between the instruction and the control system (for the particular AC) is hardwired. There is no question, “which AC”? Something similar holds for simple problems and decisions. When I decide to lift my head, I don’t need to have a my* representation since the control structures are hardwired to my decisions. Again, there is no question about who the guy is who is being instructed to move his head. And when we look at think-aloud protocols, for simple problem solving, we find few invocations of the self. However, when we turn to planning and memory, we enter the domain of deeply diachronic processes. And it’s plausible that often when we are engaged in processes that stretch back to the past or forward to the future—especially when such processes figure as components of fairly complex cognitive tasks like planning—it really is critical to keep track of the self via explicit I*/me* representations.
References


