

## Précis of *The Situated Self*

JENANN ISMAEL

*University of Arizona*

The riddle posed by the double nature of the ego certainly lies beyond [the limits of science]. On the one hand, I am a real individual man, born by a mother and destined to carrying out real and psychical acts (far too many, I may think, if boarding a subway during an hour). On the other hand, I am “vision” open to reason, a self-penetrating light, immanent sense-giving consciousness, or how ever you may call it, and as such unique. (Weyl, *Address*, 3)

### Introductory remarks

We have made huge strides on many fronts in filling in the scientific picture of the universe, both in detail and on the large scale. We have seen the integration of chemistry, biology, and physics. We also have a rapidly developing scientific understanding of the mind and the human organism. But to complete the picture, we have to show that it presents a recognizable portrait of ourselves and our conscious mental lives as we experience them from the inside. Instead of paving the way for this sort of completion, anomalies have emerged that purport to open up an unbridgeable gap between the scientific portrait of the human organism (either the low level neuroscientific description or the higher - level cognitive sciences) and the first-person experience of ourselves, an unbridgeable gap between the pumping of electrical activity in neural networks and the living, breathing conscious life we know from the inside. Instead of bridging this gap, the increasingly definite scientific understanding of the mind has only brought it into sharper focus. It is very hard to imaginatively avoid a picture in which the body is an empty shell that is brought to ‘life’ like Pinocchio by the addition of a bearer of consciousness that sees the world through its eyes. If you don’t find this picture an intuitively natural one, there are philosophical arguments that seem to make it compelling.

The difficulty hinges on the fact that we have two ways of viewing ourselves—from the inside, as we encounter ourselves in experience,

and from the outside, as represented by science—and no idea of how to fuse them into a single coherent vision. There are different ways of bringing out this divide. One can focus on the gap between the indivisible private “I” of inner experience and any material thing. Or one can focus on the gap between the quality of one’s mental life and the colourless scientific description of transfer of electricity through neural circuits. But the inability to find a way of integrating the objects of introspective awareness into a scientific world-view is an indication of a deep duality that has fuelled the swings between solipsism, idealism, materialism that have characterized the history of philosophy.<sup>1</sup>

*The Situated Self* is devoted to trying to understand how to fuse these two visions of ourselves into a single coherent picture. The central idea is that these two views coexist in the reflexive consciousness, one superimposed on top of the other, related by a gestalt shift that takes one up or down a semantic level. The book talks about the special cognitive and epistemic properties of reflexive representation, the relations between reflexive and non-reflexive representations, and the semantic level-bridging properties of self-locating thoughts, i.e., thoughts that relate reflexive and non-reflexive representations. It talks about the very special cognitive and epistemic environment created by a system reflexively representing its own interaction with environment. What follows is not so much a summary of the book as a sketch of the central themes.

### *Dualism*

The discussion in the book is centred on a collection of influential arguments for dualism. If the locus classicus for dualist argument in the western tradition is Descartes, David Chalmers is the new face of dualism. Pulling together a collection of arguments that have been kicking around in the literature and presenting them together in a powerful and distinctly modern form, his work solidified the status of phenomenal consciousness as a genuine anomaly for the scientific view of the mind. An anomaly is a problem that cannot be solved using the tools provided by the theory or world-view under examination, and its character as an anomaly—as opposed to a puzzle or problem awaiting solution, but not provably intractable—usually becomes clear only at a late stage of development.

The arguments purport to show that human selves or minds or consciousnesses are not a part of the natural world and the arguments so strong that, if they are valid, they show not only that *physics* is incomplete, but that no all-inclusive, objective representation of the

---

<sup>1</sup> It is also visible in the division between first personal and third-personal methods in psychology, and the tension between scientific and humanistic study of persons.

world is possible. No matter how wide we cast the net of representation, if the arguments are valid, they will show that the features of one's mental life of which one is reflexively aware fall outside the scope of the representation. The problem is not that you can't form an image of the world that includes your brain or your body; the problem is that the existence of your brain or body seems neither necessary nor sufficient for that of which you are immediately aware of when you reflect on your own mental life. Fill out the scientific description of your brain or body at whatever level of detail we may; there will remain an unbridgeable cognitive and epistemic divide between that description and that special ineffable, and all-important something you know from the inside.<sup>2</sup>

The picture of self as something separate from the natural order is not just a product of professional philosophy. It appears spontaneously even in the philosophically uneducated. The man on the Clapham omnibus is as hard to wean from the view of self as spirit, soul, or indivisible bearer of consciousness only contingently connected to a body as the committed Cartesian. The idea that one can outlast one's body, or that one's body might continue to exist though one's soul is absent from it, or though inhabited by someone else, is a familiar plot for movies and novels to wake up in someone else's body. The aspiring model in the TV show *Drop Dead Diva* wakes up after a car accident in the body of an overweight lawyer. Kafka's Gregor Samsa wakes up one morning to find himself in the body of a cockroach. Our sense of our own identity seems to be separable from any objectively characterizable feature or description. We think of ourselves as an invisible bit of inalienable *me*-ness that contingently experiences the world through the eyes of a particular bit of breathing matter. But the body one occupies, including the brain, is only the window through which one views the world.

These sorts of mental pictures play a more insidious role than explicitly held doctrine. They survive in a latent form even when they've been disavowed, reappearing in new guises and surfacing at crucial points in philosophical argument. They are both more interesting and more insidious than a bad theory, seeming to spring up spontaneously. Weaning oneself from them requires a patient Wittgensteinian therapy to expose them, show why we are prone to them, and how to guard against their bewitching power. Addressing the pattern of professional argument that is used to underwrite dualism is part of the goal, but the book is really an effort to get at both the psychological and philosophical roots of the tendency to think of ourselves (for each one, that is, to

---

<sup>2</sup> Chalmers focuses on the quality of mental life, but the same form of argument will show that the self isn't a material thing.

think of *her* self) as something separate from the natural order; a soul, a spirit, a primitive psychic presence separate from brain or body.<sup>3</sup> An inseparable part of the weaning process is sketching the outlines of an alternative picture, a more adequate replacement, an imaginable conception of how things might be otherwise.

The confusions that lead to dualism are layered on top of one another. They need to be teased apart and unravelled. There is a lot of probing and looking at the problem from different angles in this book. The discussion is wide-ranging. There's discussion of philosophy of mind, some philosophy of language, an injection of metaphysics, and a bit of moral psychology. There is a mix of broad stroke, big-picture sketching and fine-grained analysis. It's a matter of trying to get the big pieces of our overall picture of reality in place while recognizing that most of the details are hostage to empirical refinement and refutation. The key to the whole construction is an account of the semantic properties of reflexive representation, the self-locating representations that relate reflexive representations to their non-reflexive counterparts, and the crazy, mirrors-within-mirrors-funhouse character of a consciousness representing itself in both reflexive and objective ways.

The book is divided into three sections. The first order of business in Section I is getting rid of a view of the mind still somewhat entrenched in philosophy and replacing it with a less eclipsed image of the mind emerging from a broad coalition of philosophers, cognitive scientists, neuroscientists, and cybernetic engineers. Section II puts the new image to work to defuse arguments for dualism and diagnose the illusion that generates the dualist picture. Section III begins to fill in some holes, addressing questions about identity over time and pulling the threads together.

### **Section I: The New Image of Mind**

I use Frege in the first section to introduce the view of the mind I want to advocate. Fregean representation paradigmatically involves the manipulation of linguistically structured states whose representational content is independent of the context in which they are deployed. It portrays the mind as a detached representational system piecing together linguistically structured thoughts from a library of ideas that possess their semantic properties intrinsically and in a context-independent manner. Computation takes the form of inference and logical transformation.

---

<sup>3</sup> There is nothing inherently bad about forming mental pictures. This is a good part of how we gain understanding. But it is precisely their power that makes mental pictures potentially malignant, and there are certain ways of thinking of ourselves that are both endemic, and philosophically unhealthy.

Departures from the Fregean model in its full generality are well-under way along a number of fronts in cognitive science, but those departures have yet to penetrate mainstream philosophy and even where they have begun to penetrate, their bearing on accepted patterns of reasoning has not been appreciated.<sup>4</sup> The Fregean model works well for aspects of cognition involving manipulation of fully articulated thoughts. It notoriously does not work well, however, for indexicals or context-dependent thoughts. Examining indexicals turns out to be a fulcrum that reveals a hidden non-Fregean underbelly to thought. Indexical thoughts are the propositional analogues of coupled representations. Whereas Frege held that indexicals are special elliptical devices of convenience that can be eliminated in favour of fully articulated non-indexical thoughts, it turns out in fact that coupled representation is basic and ineliminable, indeed that fully articulated thoughts of the kind for which the Fregean model works depend on them.

To understand the non-Fregean underbelly of thought, we have to look at the information-theoretic and neurological mechanisms of cognition, including the subpersonal processes that support and sustain conscious thought. We begin not with language and logical inference, but with more basic forms of representation associated with navigation and perception involving the construction of internal models that are used to mediate interaction with the environment.<sup>5</sup> This brings out the situated, distributed nature of cognition. A mind is a machine for transforming input in the form of stimulation over sensory surfaces into behavioural response. When mediating processes get complicated enough, we start to talk about sensorimotor loops. These don't simply register information, but rearrange it into formats that are useful for behaviour management. Collections of loops can be cobbled together to form a single mind, and these in their turn can be collected under

---

<sup>4</sup> Except by a few; see, for example, Perry, J., *Knowledge, Possibility, and Consciousness*. Cambridge, Mass.: MIT 2001.

<sup>5</sup> There are first order models (e.g., Bohr's model of the atom). In these cases, the parts of the model represent parts of a physical system. There are second order models (e.g., phase spaces). In these cases, the models represent states or properties. We can construct concrete models or abstract ones; three-dimensional scale models, two dimensional figures, diagrams or maps are common. Mathematical structures described by equations or are not constrained in the same way material structures are, and can furnish a wider landscape is parsed into more useful components; parts of the phenomenal field are associated with parts of space, particular phenomenal qualities are associated with the properties of objects located in those parts. And it is their association with common parts of physical space that provides cross-modal integration (e.g., connects the sound of a chime (an element in the auditory field) with the sight of a ball hitting a bell). But these localized connections are fleeting and provisional. There is no permanent association between a part of variety of potential modelling structures.

the partial supervision of further loops. There are limitless possibilities for designing minds. At every stage in this hierarchy of loops, there is filtering, transformation, integration, and what is *given* at one level, is *constructed* or *restructured* by the levels above.<sup>6</sup> One of the things that distinguish the human mind is that we engage in self-modelling. Self-modelling is just one strategy for utilizing sensory information to produce adaptive behaviour. Self-modelling arises with the establishment of a superloop where information that is separated into different streams at the sensory surfaces is reintegrated into a unified overall model of the environment. In the context of this model, the way the world *is* is separated from one's own situation. The latter is treated as a free parameter, goals are explicitly represented, and strategies considered for reaching them.

We can get a purchase on self-modelling by thinking of a person using a map to navigate around a space in which she is located. Instead of an atlas-style reproduction of a part of space, what we really have is a fully articulated model of the world replete with causal structure and containing representations of evolving objects, living things and people with psychic histories. And instead of being embodied in an external artifact, the model is embodied in an agent's mental architecture. For a mind using an internal model to navigate, self-location takes the form of a reflexive mapping like the red dot on a map that identifies its own location in its portrait of space. A fully explicit rendering of that mapping would map the occurrent contents of consciousness into the causal structure of the world, where they would appear as a moment in the psychological history of an embodied subject. It would have the form 'this is what it is like to be so and so, at such and such a place at such and such a moment in her history'. It would do more than identify a person, time, and place; it would relate parts of the phenomenal field causally to things in public environment.<sup>7</sup>

Self-modelling is a natural development out of simpler forms of cognition. It's just one of the strategies that nature employs to allow systems to exploit the information coming in through the senses. The self-modelling superloop consolidates information and centralizes control. It doesn't supervise or regulate lower-level processes. It plays something like the role of the navigational crew on a ship. Its job is to regulate the gross motor behaviours, the sort of behaviour that is

---

<sup>6</sup> By 'constructed' we mean arranging, assembling, imposing a new order on.

<sup>7</sup> It's a little more complicated. The mapping of the phenomenal field into the objectively rendered the phenomenal field and a part of space, or between a phenomenal quality and a property of external things. These associations have to be re-established in a continuous cycle of self-location and re-location as the phenomenal field changes in response to changes in the environment and changes in our own location.

paradigmatically involved in planning and subservient to long-term collective goals. Studying the progression of content at that level in isolation is a bit like studying the minutes of meetings at the executive level for McDonalds.<sup>8</sup> There is a very great deal of bureaucratic machinery that runs without regulation by the executive level and on which the executive level depends. A proper understanding of the functioning of the mind demands a view of the whole.

We call the representation involved in self-modelling personal level representation because it is introspectively available, and the extra level of explicit representation is what opens up the space for the formation of an internal point of view. At first, this internal point of view is nothing more than the self-locating “I” that identifies a vantage point on space, but as one moves through the world, acting and observing the results of one’s actions, building up an internal storehouse of historical memories, a process begins to take hold that supports the development of dispositions, preferences, tastes, and values. The brain builds up a library of ideas that encode information about parts of the environment that form natural dynamical units—complex systems with enough internal integrity to allow them to be tracked over time. This library of ideas forms the basis for a set of concepts that are given names and become the components of Fregean thoughts and vehicles of interpersonal communication. The early glimmers of self-hood eventually turn into a fully-fledged sense of oneself as a locus of value, social agent, possessor of rights and obligations, a partner in social relationships, and so on... all the trappings of personhood.

The self-feeding supersystem that supports the development of this personal point of view is what D’Amasio and Hofstaedter and Bandura and Dennett mean when they talk about the reflexive consciousness. What this picture suggests is that there is no need for an internal substantial bearer of mental states. “I” is just a reflexive representational standpoint. What unifies the collection of thoughts and impressions that one calls one’s own is that they fall under the purview of a reflexive standpoint. The self in this portrait is a non-substantial, complex, and evolving point of view. It is a very different picture from the view of the self as a simple substance occupying and animating the body.

---

<sup>8</sup> Aspects of this mapping are what get expressed in thoughts of the form “This is what it’s like for me to see red/taste rum ice cream/feel sad in the morning...” This part of self-location is largely implicit, but it is needed to allow us to use our phenomenal states to gain information about the environment, and gets readjusted when changes in context and ambient conditions affect how properties appear phenomenally.

Personal level of representation is a quite complex achievement of lower-level processes, integrating, fusing, and synthesizing information from disparate subsystems. There are other forms of unity that we might call integrity of personality that are forged consciously in the ongoing process of self-definition. Personal integrity is an achievement rather than a given. We all have egos that are to some extent divided (there is the professional self, the playful self, the serious self, the self that wants to just play, and the self that longs for professional success). It is work to weave the pieces of one's soul into an integrated unit.

The sort of high-level cognition that the Fregean Model portrayed, consisting of the manipulation of logically structured, context-independent representational forms is coevolved with the development of language as a medium of social discourse. Fregean senses—which are the components of Fregean thoughts and the primary bearers of representational content on the Fregean Model—are connected to the world only via links that pass through more basic forms of coupled, or context-dependent representation. They are nested in a network of pre-conscious, pre-intentional relations to the world, and dependent on them for their meaning. Frege had it wrong; indexicals and context-dependent forms of representation are not eliminable ellipses for autonomous non-context-dependent referential devices. His primary bearers of meaning are utterly dependent on them for their semantic properties.

In this new image of the mind, cognition and control are both natively distributed. The emergence of a self is something that is built on top of a functional hierarchy that is self-organizing at lower-levels by the addition of a superloop of personal level representation. If the Fregean view erred in supposing that all thinking involves the manipulation of symbolically structured states that bear their meaning in a context-independent way, the view suggested by more radical trends in the movement towards distributed cognition errs in the other direction. Reacting to the lack of evidence of an informational hub and control centre in the brain and the success of decentralized models at producing some of the signs of intelligently coordinated behaviour, they propose a thoroughly distributed model. The idea is that human cognition is nothing more than sensorimotor subsystems cobbled together into coalitions producing behaviour that only has the appearance of being centrally coordinated.<sup>9</sup> Dennett has been a particularly confusing

---

<sup>9</sup> The stumbling block for such a view has always been the explanation of phenomenology. Dennett confronted the difficulty head-on, arguing that there is no such thing as phenomenology as distinct from *phenomenological reports*, confabulatory constructions intended for public relations.



figure in this movement.<sup>10</sup> In *The Situated Self*, I used him as representative of a thoroughly distributed model, and he certainly speaks sometimes as an especially radical spokesperson for such a view. The Dennett of *Consciousness Explained* and “The Origin of Selves”, in particular, likens humans to ant-colonies and insists that selves are characters in a fiction invented to facilitate the prediction of behaviour without any real correlates. But as a whole, his views are more equivocal. In other places,<sup>11</sup> he recognizes a role for a self-modelling superloop. It’s not the term he uses, but there is synthesis of sensory information, self-monitoring, and an expanding hierarchy of self-regarding attitudes out of which emerges a self. Whether there is inconsistency here or just misleading language, the real target of Dennett’s anti-realism is the Cartesian view of self as inner object running the show.<sup>12</sup> This self-monitoring superloop is not an inner object, and it doesn’t run the show. Rather than regulate the activity of subsystems (they are in charge of their internal affairs), it integrates their contents and exercises a form of selective influence that Clark dubs ‘ecological control’.<sup>13,14</sup> Gone in this picture is the conception of self as “fused control point that coordinates and regulates reflexes in discharge of action plans it itself spontaneously originates”<sup>15</sup> In its place, we have a form of top-down influence in a hierarchically organized functional system.

One of the virtues of this sort of view is that it makes sense of the evolution of selves. Selves aren’t added to the world as primitive existences; they evolve by the addition of structure to existing architecture. And it admits phenomenology. It recognizes a new level of representation in the brain at which sensory information is integrated into a three-dimensional model of the world and the introspective progression of thoughts, feelings, and experiences occurs. Neurophysiological models

<sup>10</sup> Dennett, D., *Consciousness Explained*, Penguin, 1991, “The Origin of Selves.” *Cogito* 3, 163–73.

———. “The Self as a Center of Narrative Gravity.” *In Self and Consciousness: Multiple Perspectives*. Edited by F. S. Kessel, P. M. Cole, and D. L. Johnson. Hillsdale, NJ: Lawrence Erlbaum, 1992.

<sup>11</sup> See, in particular, *Freedom Evolves*, Penguin, 2004.

<sup>12</sup> See my “Saving the Baby: Dennett on Autobiography, Agency, and the Self.” *Philosophical Psychology*, Volume 19, Number 3, June 2006, pp. 345–360, for equivocal textual evidence.

<sup>13</sup> See Clark’s discussion in “Soft Selves and Ecological Control” in *Distributed Cognition and the Will*, edited by Ross, D., Spurrett, D., Kincaid, H., and Stephens, G., MIT Press, 2007.

<sup>14</sup> Part of the response is relinquishing the strong representationalist idea that heterophenomenology is a report of the events taking place in an inner theatre, and to see it rather as constitutive or partly constitutive of the experience.

<sup>15</sup> “Introduction: Science Catches the Will”, in *Distributed Cognition and the Will*, *ibid.*, p. 5.

for how the personal level might be implemented are still too speculative to cite evidence one way or the other in its support.<sup>16</sup> Another important feature of this model of self-hood is that it makes room for moral psychology. We can make good sense of talk about degrees of self-awareness and self-control. We have the tools for an account of the self that is as complex and layered as one's *self-conception*. We open up the space for an account of the self as something that is not built into a system at the ground level, but takes shape through reflective activity in a complex, ongoing process of self-definition.

### *Dynamics*

Let's look at this through lenses that highlight the dynamical relationships. Consider a simple moving system that receives information through transducers that send waves through electric pathways resulting in movements that are designed to get the system to perform a task, or support survival in a certain kind of environment. There is feedback here; information from the system's movements is fed back to it through its transducers in ways that can affect its behaviour and produce adaptive changes in response to stimuli. A self-modelling system incorporates some extra internal dynamics. Information separated into different sensory streams is reintegrated inside the body into an internal model of the environment.<sup>17</sup> If we focus on the mental model + physical environment system, we see a tight coupling involving mutual reciprocal causal influence between the two components. You make decisions, take actions, affect the world, receive feedback from the world, incorporate it into your internal image of yourself, then the updated 'you' makes more decisions, and on it goes. A representational system whose activity is

---

<sup>16</sup> The introspective evidence for it, in my view, is incontrovertible, but these are heavily disputed issues. An especially interesting proposal due to Rick Grush shows how to achieve phenomenological unity without any centralization at the neurophysiological level. On Grush's view if we're looking to reconcile introspective phenomenology with brain science, we shouldn't be looking for a Cartesian Theatre in the brain (a specially devoted subsystem into which sensory paths feed and out of which motor pathways emerge) whose contents are the contents of consciousness, we should be looking for a higher-level virtual machine (something like the program level in a computer) stabilized out of interaction between lower level systems. As long as there are mechanisms by which information encoded at the virtual level feeds back to govern lower level behaviour, a higher-level virtual machine implementing a deliberative processes can genuinely guide behaviour. John Pollock has made a similar proposal. John L. Pollock (2008). What Am I? Virtual Machines and the Mind/Body Problem. *Philosophy and Phenomenological Research* 76 (2): 237–309.

<sup>17</sup> One's own body is represented in the model as the bit of matter that is most directly under wilful control and which mediates all other dynamical interaction with the environment.

connected in the wider domain, and that is turned back on it representing that activity, creates a feedback loop. The power of this representation-mediated feedback loop to produce new forms of flexibility in behaviour, I argued in the book, gives its dynamical *raison d'être*.

Understanding the psychodynamics of the reflexive consciousness in any detail is no easy matter. These kinds of couplings quickly outrun our ability to predict or describe, for purely mathematical reasons. The computations are literally too complex to carry out. What blossoms from this self-feeding cycle which starts with just a glimmer of reflexive awareness, is a whole internal world. If we focus more narrowly on the internal dynamics of the self-representational superloop we get a portrait of the conscious, psychic history of the agent. This sort of environment is a hothouse for the cultivation of increasingly complex representational structures and correspondingly rich phenomenology.<sup>18</sup> For a very simple example of the way that structure can propagate, think of what happens when a man at a pottery wheel inserts a thumb here, or pressure there. Programming examples are much richer. John Conway's Game of Life and its variants also provide beautiful illustrations. Feed in values of a few parameters and few simple rules (the analogue of placing a thumb here, or a finger there, and setting the wheel in motion) and you get back surprisingly and apparently complex patterns. D'Amasio and Bandura and Hofstaedter, in different ways, at different levels of explanation, have all been pushing forward our understanding of the dynamics of the reflexive consciousness. Like them, in *The Situated Self*, I argue that reflexivity is the source of much of what is special about the human mind, its fecundity and its rich capacity to form a self-feeding internal environment that functions as an incubator for new forms of representational structure. This self-feeding supersystem in the brain is what allows for the internal accumulation of pattern, pattern that is undergoing continuous organization and reorganization, all the while receiving incremental inputs from the environment.

### *Semantics*

The semantics of self-representation have long been a source of fascination for artists, logicians, and playful computer scientists. One of the hardest things to get right in the book was an account of the semantic structure of a reflexive representation. This was a crucial part of the discussion because semantic peculiarities are the source of some

---

<sup>18</sup> See D'Amasio, A., *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*, Heinemann: London, 1999 and my "Temporal Experience" in *Oxford Handbook on Time*, ed. Craig Callender, Oxford University Press forthcoming, for more detailed analyses of some of these structures.

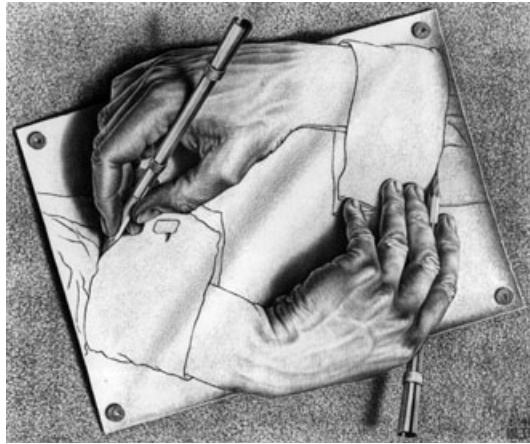
especially potent confusions about the nature of the self. Representation is a two-place intransitive, asymmetric, and irreflexive relation between a representational network (e.g., a collection of words in a language, or numbers in a coordinate system) and a domain (e.g., a physical structure or space). Starting with a domain, we can generate a hierarchy of semantic levels in which the objects at one level represent the objects at the level below. Russell's theory of types provides one example of such a hierarchy, with each language constituting a metalanguage for the language below. But there are also everyday examples of representational hierarchies, portraits of historical figures on a museum wall represent people, the numbers in a museum catalogue represent the paintings, numerals in a mathematical notation represent the numbers, and so on.

Fixed points in the representation relation can be used to interpret a representation. Knowing how to identify the fixed point (or a finite region centered on the fixed point) on a map, for example, will let us align the space of the map with the space it portrays. This is an instance of the general feature of fixed points that the mind exploits in relating itself to the world. They arise unavoidably for an agent that is embedded in the domain it is representing, and they license a form of semantic descent. The idea is that fixed points provide links between semantic levels, introducing linguistically licensed transitions from a representation of a thing to the thing itself and then back again.

We say that  $R(x,y)$  is a *fixed point* of  $R$  just in case  $x = y$ . Fixed points are a form of degeneracy. They are points at which dichotomies (in this case, the dichotomy between what is doing the representing and what is being represented) break down. Consider a sheet of paper divided into left and right sides by a line down the centre. The points on the line of bisection (i.e., at the axis relative to which left and right are defined) are fixed points; they occupy both the right and left sides of the paper, the distinction between left and right breaks down for them. Or consider, the fact that  $-40$  is a fixed point in the relation between the Celsius and Fahrenheit scales. Because  $-40$  Celsius =  $-40$  Fahrenheit, the distinction between Celsius and Fahrenheit breaks down at  $-40$ . Fixed points have special features that depend on their degeneracy. If  $R(a,b)$  is a fixed point, intersubstitution of  $a$  and  $b$  is truth preserving. That turns out to be a property that can be exploited. If you know that  $-40$  degrees Fahrenheit =  $-40$  degrees Centigrade, that can provide a transition point between the two scales. Given a metric for each of them, knowing the fixed point will allow us to calibrate them against one another.

The reflexive consciousness contains two semantic levels superimposed over one another—the presentational and the representational—

and the existence of fixed points allows us to travel by semantically valid transitions between them. There you are thinking about the world represented *in* experience, then a quick *gestalt* flip and you're thinking about experience itself, relating it to the objects represented in experience, travelling along these strange loops<sup>19</sup> like the famous Escher lithograph Drawing Hands, inside and outside of the frame.



The connections between levels made by these reflexive gestalt shifts simultaneously lend experience *representational* content and lends thoughts about the objects represented in experience *presentational* content. This gives the mind a side-on view of its own place in the world, among the objects represented in experience.<sup>20</sup> There are somewhat confusing relations of mutual inclusion between these two semantic levels. From an external perspective, me and my experience are part of the world, connected causally among the events I represent. From an internal perspective, the world is represented in experience, like a picture inside a frame. Each of these perspectives is complete by its own standards; neither is reducible to the other. The only way to get the relations between them right is to see the semantically licensed gestalt shift

---

<sup>19</sup> The term is Douglas Hofstadter's. See *I Am a Strange Loop*. NY: Basic Books, 2007. Hofstadter has done more than anyone to examine the semantic and dynamic peculiarities of reflexive representations.

<sup>20</sup> I'm using 'objects' here, indifferently to refer to anything that has a spatiotemporal location (events, bodies, systems). Objectivity of representational content in Strawson sense comes from the reflexivity; as I would put it now, the representational content of an ordinary auditory experience of, e.g., the chiming of an alarm clock, has the form "I am caused by the chiming of an alarm clock", and incorporates the side-on view of itself and its relation to the environment. The representational content of a sensory array consisting of a pattern of light, colour, smell and sound might be "I am caused by the proximal presence of a freshly baked pan of peach cobbler".

that hinges on the reflexive identity of the phenomenal field; *this* (i.e., the present contents of consciousness, including the phenomenal field) is such and such a moment in so and so's history as she peers out at the world from such and such a vantage point. Moving inside and outside the frame, travelling mentally along these *strange loops* in the mind, this is what I mean by moving in and out of the funhouse.<sup>21</sup>

One charts one's own movement through a complex and changing environment, identifying sources of danger and opportunity, acting and observing the results of one's actions, forming goals and plotting a course towards them. It's an empirical question (delicate and unsettled) how far down the phylogenetic scale self-modelling extends and what sorts of neural underpinnings underwrite it, but there's a cascade of what appear to be uniquely human cognitive capabilities that are rooted in map-making and self-modelling. Once this sort of reflexive structure is up and going, there is no end to the self-regulating hierarchy of attitudes that can be built on top of it. We start representing ourselves as situated observers of a changing environment. We then go on to represent our representational activity, forming beliefs about our representational states, and these themselves become targets of further attitudes, and so on up the hierarchy of beliefs about beliefs.

### *The Model-Theoretic Argument: Fixing Interpretation*

I said that self-locating thoughts and reflexive representations cross semantic levels, and what allows them to do so is the semantic ambiguity of fixed points.<sup>22</sup> The phenomenon of level-crossing and semantic ambiguity plays a central role in a solution of the version of the Model Theoretic Argument that I introduced as a *reductio* for a purely Fregean Model of Mind. If the Fregean Model is correct, the argument would show that reference was ungrounded. But reference *is* grounded, and the reason it is grounded is that reflexive thoughts supply semantic anchors, providing an internal grounding for reference. Let's think

---

<sup>21</sup> In external artefacts like maps, we have a special symbol that *identifies* fixed points, and the rules of use for such symbols specifically allow for semantic descent. If  $R(a,b)$  is a fixed point,  $a$  and  $b$  are intersubstitutable, and substituting  $b$  for  $a$  is a form of semantic descent. But we can just as easily embody those rules in the way it used. The difference is that in the first case, we have a user that has internalised the rules and that carries out the transitions. In the second case, we just have a dynamical system using a model of its own production to navigate a changing environment. There is no explicit representation of rules of transition, but there is behaviour in accord with it, and the reason that behaviour in accord with it works (the reason it was designed that way, or selected for) is that the transitions are semantically valid.

<sup>22</sup> The notion of *reflexive representation*, as I use it then, is a little stronger than the extensional notion of representation *of* self.

about how we establish interpretative conventions for representations like maps, i.e., artefacts designed by ourselves or others to convey information about a domain or system of interest. How do we convey the representational intent to users? We can employ language, e.g., by printing a legend in the margin written in English. If we're located so that we can see the landscape represented on the map, we can use ostension. An agent using the map who sees the landscape in front of her can point to parts of the landscape and point to parts of the map and say 'this represents that'. We can affix labels drawn from the map to parts of space. What do all of these ways of establishing interpretive conventions have in common? The use of a medium—e.g., language or vision—external to the one being interpreted, in which the object being interpreted and its intended domain are represented jointly and related to one another. What makes the interpretation of *mental* representation problematic from the point of view of establishing interpretive conventions is that in that case Putnam asserts that there are no external media that the mind can use to get a side-on view of the relationship between itself and the world. That is the point, as I understand it, of his insistence that one can't leap outside of one's own head and hook his thoughts directly up with things. The mind is confined by definition to using some form of mental representation to pick out objects. When a child learns a word, he says, the child doesn't learn to associate the world with a thing, he learns to associate it with other mental phenomena (images, smells, etc.) that become united by being regarded as sharing reference.

So what does the New Image of the Mind (the one endorsed in *The Situated Self*) say about the interpretation of mental representations? The standard naturalist response to Putnam goes something like this: the mind's inability to get a side-on view of its own relationship to the world is neither here nor there. All that is required of any representational medium is that it be coordinated with its subject matter in a manner that allows it to act in the contexts in which it is employed as a vehicle for information about that subject matter. We need interpretive conventions for public artefacts—words, maps, and symbols on bathroom doors—only because they don't have a natural role in our lives, and so conventions are needed to govern how they are to be used. But we don't need to establish interpretive conventions to coordinate mental images and other forms of mental representation with what they represent any more than the frog needs to establish interpretive conventions to coordinate fly images with flies. And if we don't need interpretive conventions to coordinate mental representations with the world, we don't need an internal grounding for their interpretation. What makes it the case that A represents B for the frog's image of the

fly, the child's image of its mother, the dog's idea of its master is a certain kind of causal-informational link, between A and B (together, perhaps, with some ancillary facts, e.g., facts about the biological function of A or the history of A, or whether learning can modify the connection between A and B). These are by their nature external to the mind. Our representational intent plays no role in establishing them, and it doesn't matter that we don't have an illuminating way of representing them to ourselves.

The naturalist who defends a broadly causal-informational account of mental representation rejects any requirement on thought except that it be appropriately coordinated with the world (where coordination is made out in purely causal-informational, or dynamical terms, without employing intentional or semantic notions), accepts that the mind doesn't have a side-on perspective of its own relations to the world, but holds that thought is none the worse for it. The frog mind may not have a side-on perspective of the relation between its internal image of the fly and the fly itself, the lion mind may not have a side-on perspective of the relation between the smells and sounds and visual manifestations of deer, but that doesn't keep either of them from sighting, tracking, and responding in all of the right ways to their presence. There is no reason, the response holds, for thinking that the human mind is different from that of other animals in this respect. It is part of a complex machine coordinated with the environment by causal-informational links designed by evolutionary history to keep it moving in the right ways in response to the right sorts of stimuli.<sup>23</sup>

Reflexivity goes beyond the standard naturalist story and adds something quite important. It agrees with the naturalist that what matters for understanding the relations between mental representations and the world is coordination, and coordination doesn't require an internal

---

<sup>23</sup> The substance of a causal-informational account goes into specification of the relevant sort of link and is guided by an explicit characterization of the relation that coincides extensionally with an intuitive assortment of examples and counterexamples of representation. The hard part of a naturalistic story is finding relations discriminating enough to rule out extensionally equivalent interpretations of ideas. The most popular approach, associated with Dretske and Fodor, has used concepts of physical correlation with some kinship to those employed in information theory. Teleosemantics is a form of naturalistic semantics that uses a concept of biological function developed in the mid 1980s with work by Millikan and Papineau. I have here ignored the class of metaphysical responses to the Model Theoretic Argument following David Lewis' lead, because they're a little outside the scope. But see "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61 (1983): 343–77. For our purposes, it matters only that they are externalist accounts. The facts that fix the intended interpretation of thought and language are external to the mind; and they don't recognize an internal grounding for interpretation.



grounding for thought. But it disagrees in two respects: it denies that we don't have an internal grounding for reference. Self-locating thoughts of the form 'here is ...' and 'now is...' serve to ground reference of spatiotemporal vocabulary in the 'here-now' of the present thoughts. Self-locating thoughts of the form 'this is what it's like to ...' ground reference in phenomenal qualities that are causally related to elements in the external landscape. And it denies that we do not in fact introduce and make heavy use of representational conventions in mental representation. The mind does have a form of ostension in its bag of tricks—*viz.*, reflexivity—that allows it to explicitly introduce new forms of representation and relate them to the world, and to consolidate and clean up the loose, redundant forms of representation that mother has provided it with. And the way that the mind employs this trick to introduce new, conventional forms of representation is not incidental. I argued in *The Situated Self* that it goes right to the heart of what is special about the human mind, its creative ability to construct explicit models of the environment, to revise and refine them in light of new information, to entertain new theories, and develop new languages. Because reflexively identified elements are connected both in the space of representations and in the space of things, they can serve as points of reference with respect to which we locate other things. We can use them to set up stable systems of objective coordinates that cover the whole space grounded by reflexively identified times, places, and properties, provided only that we have enough of them and that they're well-situated.<sup>24</sup> In an n-dimensional space, provided there are no unbroken symmetries, it takes in principle only n points of reference to cover the space.<sup>25</sup>

It's an important virtue of the naturalist story that it makes human cognition continuous with animals lower on the phylogenetic scale. But it's a problem with the standard naturalist story that it leaves the sharp, apparent discontinuities in capacities unexplained. The exploitation of

---

<sup>24</sup> Using 'space' here to refer to any set of elements with a relation defined over them. A system of coordinates *covers* a space just in case each element in the space is assigned a (unique) set of coordinates. In a well-organized system of coordinates, these names will themselves encode a lot of information about relations of points to one another.

<sup>25</sup> An inscription of the form "I [this inscription] am located at precisely <35 degrees, 24 degrees>" will not only relate the point at which the inscription occurs to the coordinate description "<35 degrees, 24 degrees>", it will relate Iceland to its coordinate description, and the Eiffel tower to its coordinate description, and the North Pole to its coordinate description. "I [this utterance] occur at time midnight august 18, 2007" will not only relate its own utterance occurs to the temporal description "August 18, 2007", it will relate Elvis' death and Mozart's birth to their temporal descriptions.

conventional representation and reflexivity is part of what distinguishes human cognition from animals and it explains those discontinuities. All cognition of any type depends on background conditions and employs rules of transition that transform and combine representational states. What's special about human cognition is our ability to bring the background conditions and rules of transition into the representational content and make them objects of thought. That ability to think about thinking, and in particular to think about the rules of thinking and the external conditions that support representational content, lends thought a very special kind of adaptability.

## Section II: Defusing Arguments for Dualism

Section II addresses Chalmers' arguments for dualism, by showing how reflexive awareness and an external or scientific description of the mind can coincide extensionally despite the apparently unbridgeable cognitive and epistemic void between them. I recognize that there is understandable *ennui* over the 'hard' problem of consciousness, but I revisit them without apology, because there is still no clearly adequate solution, still new things to be said, and still a great deal to be learned from it. Chalmers' arguments focus on phenomenal consciousness, the introspectively grasped quality of mental life. To get the right account of phenomenal properties, we need to generalize the notion of a fixed point to a reflexive surface.<sup>26</sup> I said, above, that the distinction between left and right sides of a sheet of paper breaks down at the line of bisection. That is because every point on that line is a fixed point between the left and right sides of the surface. If we consider a sphere bisected by a plane into two equal volumes, every point on the plane of bisection is a fixed point in the relation between the left and right sides of the sphere.

Consider a room in which one wall is covered by a mirror. Suppose the room contains a table, a few chairs, and a dish of apples all within view of the mirror so that the mirror contains reflected images of these things. There is all the difference in the world between the things and their images; the table and chairs in the room are three-dimensional and made of wood. The tables and chairs in the image are two-dimensional and made of glass. What makes the latter images *of* the former is that the former represent the latter under the natural interpretation of the images in the mirror. There is a point-by-point mapping which associates three-dimensional volumes in the room with regions on the mirror's surface. These are, for the most part, disjoint,

---

<sup>26</sup> This is not a term I used in the book, but I think it's helpful.

except for the surface of the mirror itself, which occupies the interface between the two. Each point on that surface is a fixed point in the relation between the mirror and the room, and can serve as a point of reference in relating objects in the image to objects in the room. It has precisely the sort of ambiguous status that the phenomenal field has in thought. The phenomenal field is the reflexive surface of thought.<sup>27</sup> It stands at the interface between the external world represented in thought and the internal world of representations and has the special semantic duality of fixed points. These are not two separate spaces, closed with respect to one another and related only externally by the representation relation  $R(x,y)$ .  $R(x,y)$  has fixed points and the semantic duality of these points is exploited to relate experience to the world represented in that experience in the Mother of all self-locating thoughts: “this is what the world seems like to me in the here-now; this is what it’s like to be such and such an embodied individual experiencing the world from such and such from such and such a vantage point at such and such a point in time.” A reflexive consciousness has an Escheresque structure with two semantic levels (the reflexive superimposed over the non-reflexive) related by thoughts of this form.<sup>28</sup>

William James makes some suggestive comments in the same vein:

The puzzle of how the one identical room can be in two places is at bottom just the puzzle of how one identical point can be on two lines. It can, if it be situated at their intersection; and similarly, if the ‘pure experience’ of the room were a place of intersection of two processes, which connected it with different groups of associates respectively, it could be counted twice over, as belonging to either group, and spoken of loosely as existing in two places, although it would remain all the time a numerically single thing.<sup>29</sup>

All of this goes toward an explanation of why “this is what it’s like to...” thoughts are informative and allow us to sidestep the conclusion of perhaps the most difficult and powerfully persuasive contemporary

---

<sup>27</sup> The properties of the surface are related in to the properties of the objects it portrays in a way that is mediated by physical processes and reproduces to a fairly faithful approximation the relation between phenomenal properties and the external properties they are used to track.

<sup>28</sup> See my “Doublemindedness; a model for a dual-content cognitive architecture.” *Psyche*. July, 2006, 1–11.

<sup>29</sup> James, W., ‘Does Consciousness Exist?’ <http://psychclassics.yorku.ca/James/consciousness.htm>.

argument for dualism: the Knowledge Argument.<sup>30</sup> These thoughts relate a reflexively presented property to a description and assert extensional equivalence. Reflexive, or presentational knowledge is unmediated, non-discursive. It is a matter of showing rather than saying, exhibition rather than denotation. The resulting view doesn't do anything to remove the mystery of consciousness. It preserves the ineffable character of phenomenal 'knowledge', but exposes misplaced expectations about what ontological integration of the sort required for the completeness of physics demands. In the book, I say why a Chalmers-style positive proposal for a science of consciousness is misplaced, and recommend a Wittgensteinian quietism about 'what phenomenal properties are'. The right account of the structure of concepts, on my view, sees all descriptive concepts as rooted extensionally in phenomenal concepts [for there to be something it is like for me to be in S, means that S has a quality that is on display in thought] and exposes the mistaken assumptions that lead to the expectation of an informative scientific account of the nature of phenomenal consciousness. We need *some* properties that can be displayed to act as semantic anchors for the interpretation of descriptive vocabulary. Phenomenal properties play that role.

There is a catch, however, that is raised by the epistemic possibility of inverted spectra, which serves as the basis for another of Chalmers' arguments for dualism. Reflexive, or presentational knowledge, the sort of knowledge expressed in 'this is what it is like to...' thoughts is incommunicable interpersonally because the presentational domains of different interlocutors don't overlap. There are no joint experiences, which can provide a common frame of reference against which to calibrate the qualities on display in experience. Language has a dual life, in public and in private. Linguistic representation in thought emerged together with language as a public phenomenon, and is dependent on its public use as a medium for exchanging information about a common environment. On the view defended in *The Situated Self*, qualitative content comes from its private role, association with phenomenal states in 'this is what it's like to...' thoughts. This allows the incommunicability of phenomenal knowledge, without denying the possibility of cognitively significant phenomenal thought. The resulting picture is a non-reductive naturalism that integrates us into the natural order, but preserves something quite central to the dualists (the ineffability of phenomenal knowledge, the possibility of spectral inversions).<sup>31</sup> This is not a concession. The recognition that in epistemic and semantic respects,

---

<sup>30</sup> Chalmers presents the argument, but the original version is due to Frank Jackson.

<sup>31</sup> Notice that in Section I, I also preserve something quite central to the Cartesian style of dualism, viz., the centralization and the unity of consciousness.

physical thought rests on the phenomenal is simply a form of empiricism, one that is entirely compatible with ontological monism.<sup>32</sup>

Chalmers introduced a division of questions about the mind into ‘hard’ and ‘easy’ ones. The easy questions are questions that can be addressed scientifically. These include questions about the functional structure of the mind, the source of specific cognitive capacities, the neurophysiological basis of language, and so on. The hard questions, by contrast, don’t permit of an informative scientific answer. These include questions about the nature of phenomenal consciousness. The ‘easy’/‘hard’ terminology may be a misnomer. In one sense, the so-called hard questions are very easy on my account. They permit a trivial, but uninformative answer. They can be answered by showing, but not saying. If I ask myself what phenomenal consciousness is, or what it is like, for example, to see red, I give a reflexive answer. I get myself into a situation in which an instance is present to the mind and I think “It’s like *this*”. A reflexive account resolves the ‘hard’ problems while leaving a great mass of complex and fascinating questions both of theory and detail that will however yield to ordinary scientific and analytical investigation. It removes the *in principle* stumbling blocks that dualist arguments were supposed to expose, while leaving all of the interesting and *doable* stuff.

Turning from philosophical arguments to a more elusive target, the picture of self as a primitive locus of mental life has a very strong grip on the imagination. I think we all have a strong sense of the internal simplicity of the self as psychic presence and a tendency to think of our selves as separable from not just any, but *all* of our properties. I suggest in the last part of Section III that the psychological roots of this idea may lie in something as simple as hasty generalization of the example of self-location on a map, where the user thinks of the self-locating red dot as a representation of himself. This suggests the need for an object that furnishes a referent for the “I” in self-locating thoughts. But the “I” can’t be a body, because I can have a complete description of the locations and dispositions of all of the bodies in the world, it seems, and still not know where I am, because I don’t know which body is mine. So I must be something that lives *inside* a body, something separable from it in thought, only contingently associated with it: a pure subject, bearer of experience, author of judgment and volition. The reflexive account of self-location defuses this line of reasoning by getting rid of the user as an ineliminable intermediary between a representation and the world, treating the agent as the physical embodiment of a

---

<sup>32</sup> And a form of empiricism that can be defended (though not here) against Sellars’ criticisms of the Myth of the Given.

self-model, and the 'self' in 'self-representation' as not a name for an object, but as indicating the reflexive nature of the locating relation, like the self in 'self-supporting' or 'self-undermining'. Gone is the self as pure ego. What we have left is an active mind that conceives of itself as a reflexive surface on which the world, including its own actions in the world, project images of themselves. Where Descartes inflates the self to the status of an inner object, a reflexive account absorbs it into self-representation, leaving us with an internal point of view as part of the representational content of a self-image, but no inner object occupying that point of view.

### Section III: The Non-substantial Self

So what, according to *The Situated Self*, is the "anatomically invisible, terribly murky thing called I"? All that the grammar tells us is that I am the proper subject of a reflexive act. That leaves 'I' as equivocal as 'here' or 'now'. We draw the boundaries differently in different contexts. The minimal self is the here-now of the individual thought. By embedding this in a conception of personal history we get the temporally extended I of the autobiographical self. By forming a richer, hierarchically structured conception of self as moral and social agent, we can get more complex self-regarding attitudes that single out some core of commitments and self-defining character traits.<sup>33</sup> The temporally wide self, the rational self, the virtuous self; these are all products of unification, forging reflexive points of view that span and integrate the contents of states that are distributed across time, or across different cognitive subsystems, or belong to different compartments of the ego. They are all nested within the tangled hierarchy of the mind. Each of these has its own brand of unity. The only one I discuss in detail is the unity of the autobiographical self, arguing on what might be termed 'grammatical' grounds for a Pure Lockean solution to the problem of the self over time.<sup>34</sup>

The sort of unity that is relevant to Descartes' 'I' (i.e., to a thinking thing, to an intentional subject) is different from this, and also from the sort of unity that is relevant to a material object. It has to do with the unity of an intentional standpoint. This is discussed here only in passing, but I now see it as a very important element in a complete naturalistic account of human selves. It is not mereological in character at all, but formal. Once we've cleared away the philosophical and

---

<sup>33</sup> There is also room, as Millgram notes in his comments, for smaller, compartmentalized subselves (a professional self, a maternal self, an intellectual self). We have uses for both thin and more complex notions of self-hood

<sup>34</sup> The view is very close to David Velleman's in "Self to Self." *The Philosophical Review* 105, no. 1 (1996): 39–76, and owes a great deal to his insights.

psychological demons that lead to dualism, the positive research program that remains is a reconstruction of the phenomenology of living, the full felt quality of life for a normal human, the cycle of presentation and representation driven by a stream of sensory information and memory, guiding action and experiencing the results of its movements. The ontological project leaves in its wake phenomenology, and moral psychology, and practical ethics.

### *Perspective*

I've foregrounded the idea of the reflexive consciousness and how a proper understanding of reflexivity can defuse influential arguments for dualism, but the book also provides a bit of useful regimentation on related matters. The logic of perspective is something we all have an intuitive grip on from experience with the spatial example. The mathematical treatment has been worked out quite well in physics, and injection of that apparatus can go a long way to systematizing talk of perspective in the philosophical literature and clarifying why certain kinds of information 'go missing' when one adopts a non-perspectival mode of representation. The native position of the embedded agent is a frame-dependent mode of representation. We construct a (relatively) frame-independent mode of representation by taking what is invariant under transformations between frames, and we use the frame-independent representation as a way of coordinating across frames. We do this with an increasingly wide set of transformations, pushing towards a more and more abstract representation of the world, invariant under an ever wider class of transformations. Frame independence is always a relative matter; it is relative to a class of transformations. What does one leave out when one retreats to a model of representation that is invariant under transformations between temporal perspectives? Differences between now and later, past and future, soon and a long time hence. These are reconstructed as relations to moments of time. What does one leave out when one retreats to a model of representation that is invariant under transformations between spatial perspectives? Differences between near and far, up and down, here and there. These are reconstructed as relations to oriented regions in space. More controversially, what does one leave out when one retreats to a mode of representation that is invariant under transformations between temporal perspectives? Change and movement. These are reconstructed as relations between parts of a world-line.

The epistemology of science is continuous with the epistemology of common sense realism about the world. We see ourselves as working our way outward from an embedded (or internal) perspective, forming

increasingly frame-independent view of the world, using increasingly invariant concepts, but in a manner that leaves them rooted extensionally in places and objects that can be ostended and in properties present/on display/ exemplified in the context of use. Physics *aims at* a mode of representation that is invariant under transformations between perspectives—at the very least, spatial, temporal and personal perspectives—and it is not surprising that it leaves out information of significance from these perspectives.

I want to emphasize in closing the précis and turning to the comments of symposiasts that *The Situated Self* does not offer a reductive account of consciousness. I assume that the level of representation that is characterized by reflexive awareness is coincident in extension with consciousness, but I don't think that consciousness just *is* reflexive awareness.<sup>35</sup> There are many examples of systems that are reflexively self-aware in the sense discussed in *The Situated Self* (i.e., governed by a self-modelling superloop), but that are not conscious. My strategy is rather to deflate misplaced expectations that science could or would or should be held responsible to providing a reduction. Consciousness, as such, is the very paradigm of something that can be shown, but not informatively described.<sup>36</sup>

Here are a few lines from Walt Whitman's, "When I Heard the Learn'd Astronomer", which capture one of the central take home lessons of *The Situated Self*.

When I heard the learn'd astronomer,

When the proofs, the figures, were ranged in columns before me,

When I was shown the charts and diagrams, to add, divide, and measure them,

When I sitting heard the astronomer where he lectured with much applause in the lecture-room,

How soon unaccountable I became tired and sick,

---

<sup>35</sup> See Kriegel, and authors in Kriegel volume for discussion of self-representational accounts. Self-representational accounts are seen by some as modifications of the higher order thought view, and presented as constitutive accounts of consciousness.

<sup>36</sup> I have reservations that there is a general quality that all conscious states share, as opposed to a collection of particular qualities that can't be communicated or compared across minds.



Till rising and gliding out I wander'd off by myself,  
In the mystical moist night-air, and from time to time,  
Look'd up in perfect silence at the stars.

The difference between presentation and representation, and the fact that the distinction breaks down for reflexive representations, is right at the heart of *The Situated Self*. Description in an objective vocabulary is representation, and representation is not *presentation*. It is no substitute for the thing itself. The whole point of representation (and what gives the answer to Rebecca West's of-repeated quip about why we would want a *copy* of the world: 'isn't one of the damn things enough?') is that representations do not have (exemplify, possess) properties of the things they represent.

#### *Loose ends*

What precedes is not so much a summary as some thoughts about the central themes, as I see them now, and that means there has been some filtering and transformation. There is some related recent work that sheds light on some of the themes in *The Situated Self* and is worth mentioning. I departed in *The Situated Self* from the standard practice by speaking of mental representation in terms of the construction and manipulation of mental models. A model is a concrete or abstract representation of a system that acts as a representational proxy that can be viewed, manipulated, or used for prediction or testing. It has parts that correspond to parts of what it represents, and in dynamic examples, it evolves and acts in the same way as what it represents under intervention. A *self*-model falls within its own representational scope and has a reflexive element—like the red dot on a map—that identifies its own place in the landscape. Models have a compositional structure and dynamics that is different from the grammar and logic of propositional representation. Cognitive scientists are recognizing that much human cognition takes the form of modelling. Working with models rather than language gives me a rich source of visualizable examples that capture the structure of a reflexive representation and allowed me to focus on the aspects of cognition that I was interested in and avoid some of the complications introduced by language. Someone using a map to navigate a terrain is implementing a self-representational loop of exactly the sort that the mind implements in processing sensory information through an internal model of a spatially ordered world. There is still, however, an unspoken orthodoxy that leads philosophers to think of

mental representation in propositional terms. In *Models and Cognition*, Jonathan Waskan makes differences between modelling and propositional thought his focus. He challenges the orthodoxy, arguing that the construction and manipulation of cognitive counterparts to scale models is what most clearly differentiates human cognition from other creatures. The idea that modeling (and in particular self-modeling) is a characteristically human development, and one that is at the root of a matrix of special cognitive abilities is one of the central convictions of *The Situated Self*. Waskan's book is a wealth of fascinating detail and scientific data that one hopes could lead to a reexamination of the orthodoxy. I continue to think that something special and unprecedented happens when humans develop language. But that doesn't detract from the importance of non-linguistic forms of cognition.

The epistemology in Robert Stalnaker's *Our Knowledge of the Internal World* is very close in spirit to the one in *The Situated Self*. It shares with *The Situated Self* the idea that epistemology starts from the inside, with phenomenal knowledge, and works its way out simultaneously to a conception of the objective landscape and of ourselves as epistemic agents in that landscape. Stalnaker also discerns a connection between self-location and knowledge of one's own phenomenal experience. The details and the framework are different and the analogy with self-location is developed differently but the epistemology at the heart of the book is closely akin to *The Situated Self*.

Francois Recanati's *Perspectival Thought: A Plea for Moderate Relativism*<sup>37</sup> develops semantics for context dependent thought that is of a piece with that in *The Situated Self*. He recognizes and formalizes the semantic contribution of what I called the non-Fregean underbelly of thought (and does much else besides; the book is full of interesting semantic exploration). He is one of the good-guys, having made the Copernican shift that sees context-dependent forms of coupled representation as basic, and fully articulated Fregean thought as a kind of representational satellite that depends for its lifeblood on them.

---

<sup>37</sup> Oxford University Press, 2007.