Separation and Engagement: From Duplex Vision to the Achievement of Self-Consciousness

Sara Saab

1: Introduction

The world we observe and the world we act in are one and the same world, and crucially so. Dualisms are rife in cognitive science, and superficially, this paper may seem to be an exercise in splitting things in two. Neither brain structures nor cognitive aptitudes nor even modes of awareness will appear safe. Yet the core insights of this paper are guided by the above maxim of metaphysical monism, and serve to reinforce it. They will concern not differences and oppositions, but rather cooperation and complicity. To employ a prime metaphor elaborated by Brian Smith (1996), this is one story of the single dance we perform in our world, and how the two main skills we bring to bear therein – and to dig deeper, *how their neural foundations* – result in our important achievements as thinking creatures. Though it may traffic in popular dualisms, this is a tale of negotiation and mutual enrichment.

Monisms and dualities notwithstanding, there is also a three-step hierarchy that the following argument will scale. I will begin in the trenches of neuroscience and the psychology of visual perception. This will motivate a discussion of two skills available to perceiving creatures, skills that I will subsequently claim to be central to a rather sophisticated form of self-consciousness. The main charge of this paper may be conceptualised as the task of showing that these three domains, of visual perception, of vital ways of interacting with the world, and of self-consciousness, are importantly interconnected.

In what follows, I will build the case for two broad realms of embodied, embedded cognitive capability, referred to as *separation* and *engagement* in honour of Smith's (1996) usage, being directly enabled by the functions characterising ventral stream and dorsal stream visual processing, respectively. Separation and engagement are our twin abilities to represent what is at a spatial or temporal distance from our local and present surroundings on the one hand, and to interact with our immediate, available environments, on the other. I will commence by exploring the perceptual performances made possible by the ventral and dorsal streams of the visual system. Based on empirical findings and theoretical analysis, I will draw out the relationship between ventral processing and a modest form of separation, and indeed, between dorsal processing, ecological perception, and engagement.

This strictly segregated dialectic will soon begin to seem artificial. It will therefore be synthesised by way of arguments for the necessity of dorsally mediated

processing for full-fledged separation, and the metaphysically indispensable position of the functions of ventral processing in engagement as we know it. Finally, I will attempt to show that full-fledged separation is what catalyses the transition from being an aware subject to being an object of self-awareness. Yet full-fledged separation cannot exist in the absence of an engaged, active life and the neural processing that supports it. Therefore, my claim will be that a sophisticated kind of self-consciousness can be traced back to the functions of the two visual streams via the interlaced achievements of separation and engagement. Prior to my closing remarks, I will advance some clarifications of this thesis and field objections as to its implications for the nature of cognition.

2: The Visual Brain

The mammalian brain, many researchers agree, processes visual scenes along two major neural pathways (Goodale & Milner, 2004; Carruthers, 2005; Norman, 2002). In fact, a variety of organisms, mammalian or not, display a division of labour in visual processing. Frogs have one system for dodging obstacles, one for controlling tongue-snaps in the direction of food, and at least three other visuomotor modules (Goodale & Milner, 2004).

What is interesting about the division of visual processing in mammals – what is interesting about humans – is exactly what our two major visual streams function to do. In humans, vision begins with light impinging on the retinal surface. There, light wavelengths are transduced into neural activation and signals are sent up the optic nerve, via a pit-stop at the lateral geniculate nucleus roughly midway along the length of the brain, to the primary visual cortex (V1). This area in the occipital lobe at the back of the brain is a sorting hub for the all-but-raw visual information. A coarse sketch of such properties of the visible tableau as colour, edges, textures, and motion is pieced together before it is relayed back toward more frontal brain areas for further processing (Goodale & Westwood, 2004; Clark, 2004).

The two major routes for information originating in V1 form an acute angle, one route climbing *dorsally* towards the crown of the head, particularly to the back of the parietal lobe, and the other diving *ventrally* towards the area between the ears, to the lower regions of the temporal lobe. These two paths are called the dorsal and ventral visual streams, respectively. It is largely these neural pathways that make our

experiences of and responses to our visual world possible. There are additional direct connections between the retina and dorsal areas via evolutionarily more ancient midbrain structures. These visual structures in the human subcortex happen to be the dominant visual centres in non-mammalian vertebrates. For example, it is an analogue of this area which controls the aforementioned tongue-snapping motion in frogs (Goodale & Milner, 2004).

A huge corpus of work in vision science has accumulated around the so-called *dual visual system* consisting of the V1-to-ventral pathway and the V1-to-dorsal pathway. Much of this work has sought to demonstrate that these streams deserve to be differentiated functionally as well as anatomically (cf. Goodale & Milner, 2004; Westwood, Chapman, & Roy, 2000; Grill-Spector, 2003).[1] Insights from neural pathologies have led to a functional characterisation of the ventral and dorsal streams that has since received a great deal of empirical support.

The first relevant disorder, *optic ataxia*, is a condition following lesions along the parietal dorsal stream. Sufferers have no trouble at all recognising, labelling, reproducing, or remembering items that populate their visual scenes. Asked to act upon those items, however – to reach for a comb or open a door – the movements of optic ataxics are exaggerated and inappropriately scaled.[2] Conversely, a disorder called *visual form agnosia* spares the ability to execute neat and appropriate actions within a visible environment, but strips patients of the capacity to name, draw, or otherwise indicate what things they can see (Westwood, Chapman, & Roy, 2000). For visual form agnosics, who have damage to portions of their ventral streams, achievable actions grossly out-span any demonstrable knowledge of the contents of the visual scene (Norman, 2002).

The functions lost in optic ataxia and visual form agnosia respectively – the functions subserved by the two visual streams – can therefore be said to have been *doubly dissociated* from one another in empirical settings (Goodale & Milner, 2004). The usual interpretation of double dissociation, wherein each function can be shown

---

[1] Inversely, holding a functional – not architectural – divide to be basic, some researchers have sought to attribute functions to medial regions between the parietal and temporal lobes so as to blur the purported strict anatomical segregation between streams (Gallese, 2007).

[2] A recent review by Rossetti and colleagues disputes the dominant characterisation of optic ataxia, hypothesising that the disorder is a deficit of on-line control which becomes significant only when action programming processes are taxed, as when targets are in the periphery. The import of this alternative treatment remains to be determined (Rossetti, Pisella, & Vighetto, 2003).

to exist though the other function is lost, is that the underlying mechanisms which support those functions are in some sense independent (Bedford, 1997). But what exactly are the functions carried out independently by the ventral and dorsal streams? Originally, it was thought that the two pathways engaged in object discrimination and spatial localisation, respectively. This led to their being labelled the *what* and *where* pathways. Nowadays, the dorsal stream is thought to enable visually guided action, like reaching, ducking, and navigation. For this reason, it is now known as the *how*, not *where*, stream (Jeannerod & Jacob, 2005).

The visual world according to the dorsal stream is a topography of information to facilitate motor interaction with a scene, including a sequence of *egocentric invariants* – measures stable enough with respect to the eyes or other body parts to be useful for visuomotor performance (Norman, 2002; Cohen & Andersen, 2002). These invariants are calculated quickly and with astounding accuracy. They are *absolute* measures, which is to say that they do not represent the parts of the scene to which action will be directed in relation to other parts of the scene; that could result in size or shape or orientation illusions. After all, with respect to the height of professional basketball players, the net might seem about a metre out of reach, but if I attempted a slam-dunk based on this information, my movements would be utterly inept.

Nearer the base of the brain, a different world is pieced together. It is chiefly in the ventral stream that a scene populated with individual objects, as opposed to a canvas of spatiotemporal properties, is constructed. For example, neurons selectively sensitive to houses and faces can be found in the ventral regions of modern primate brains (Grill-Spector, 2003). It is useful to think of ventral areas as processing *vision-for-judgement*, in other words, as providing a visual landscape of information to answer questions of various cognitive complexities: what is that thing? how heavy is it? will it fit? is it mine? Another perspective is one advanced by Jeannerod and Jacob (2005); the ventral stream, they claim, is occupied with the task of *perceptual selection*:

> The selection phase consists in both segregating a complex visual array into several separable objects and in attributing to each separate object its own set of appropriate visual attributes (this is the so-called 'binding' problem). Usually the color and texture of an object will be highly relevant to its perceptual selection from a set of neighboring objects. Segregation and binding require that the relative spatial locations of different objects in a visual array be coded (p. 303)

As Jeannerod and Jacob point out, ventral processing is *relational* instead of absolute. The ventral stream happily compares the height of the basketball net to the height of the players – only then can it partition the scene into items moving relative to the stationary background, reaching about this high off the ground, and on the other hand, an item that begins to fill space a little bit further up (Jeannerod & Jacob, 2005). Ventral processing is slower than dorsal processing. It is furthermore *allocentric*, with objects being selected in a perspective-neutral way (Schenk, 2006). Perspective-neutrality is achieved by encoding enduring properties centred on spatial locations in the world, producing measures which are not directly altered by every eye saccade and head movement (Norman, 2002). Allocentric processing, in turn, facilitates the tasks of object recognition and identification.

Object recognition and identification may, incidentally, seem strange things to add free of charge to the repertoire of ventral processing. It is one thing to see objects, and quite another to see *which* objects there are. Plausibly, the latter could involve non-visual neural systems. The specific topic of recognition and identification will be treated later. Nonetheless, a general methodological concern remains. How do we know when we are conflating the processing of one brain region with that of other regions? And to turn the tables, is it valid to theoretically isolate ventral or dorsal processes from one another, and from the remainder of a network with which they are massively interconnected, especially considering that some aspects of visual processing occur in more primitive brain areas like the superior colliculus, while others take place in higher cognitive areas like prefrontal cortex (Goodale & Milner, 2004; Jeannerod, 2006)? My own explorations will involve prising apart ventral and dorsal functions. Is this justified?

In best appeasing the competing forces of understanding brain area contributions and respecting neurofunctional interdependence, a good starting point is the robustness of the cortical brain. Brains are surprisingly good at withstanding local damage. Even functions which usually cooperate in cognition, like having kinaesthetic sensations in a limb and knowing the limb to be one's own, can be dissociated. Dissociation makes it sensible to analyse the brain by evoking somewhat localised, independent functions which are eventually integrated instead of insisting on black-box holism. It also provides a heuristic for judging whether distinct processes are being conflated. If ventrally lesioned subjects can identify objects – based on verbal descriptions, perhaps – and vice versa, then vision and object

identification ought to be treated as importantly distinct. Such functional segregation and anatomical segmentation should not be overdone, however. Specifically, empirical demonstrations of dissociation are on much securer ground than theoretical stipulations that one area does not depend on the operation of another. Since I want each visual system in particular and not just the brain in general to do some work in my account, I will engage in the latter ill-advised project to put maximal pressure on my own arguments for each visual stream supporting special cognitive skills (particularly in upcoming discussions of creatures I will call Vee and Dai).

Fortunately, this approach is only temporary. I will eventually discuss higher-level skills enabled by sets of functions which it would be foolish to imagine existing without interweaving contributions from several parts of the brain. As these skills come to the fore, I will abandon the strategy of underemphasising functional cooperation in order to give proper weight to how rich and vital those skills are. In the meantime, the main danger my temporary strategy poses is that attempts to explain which neural processes enable which skills, and which skills enable which sorts of consciousness, could turn out to be futile because those processes, skills, or types of consciousness are direct outcomes of neurofunctional interconnections which I happen to have overlooked. Towards the end of the paper I will address this issue, advancing reasons to believe that the order of developmental sophistication is as I claim it to be; it is raw visual abilities which bring on cognitive aptitudes, which in turn promote new modes of awareness.

That said, we may properly return to the world of the visual system, for a few points remain to be made. To recap, the dorsal stream speedily processes visual information from V1 and the midbrain into egocentric, absolute metrics which facilitate enactive interaction with the visual world. The ventral stream uses information from V1 to construct a world of objects. It employs allocentric and relational measures to arrange bunches of properties into items. Much more will be said about objects in what follows, and the issues of recognition and identification will be revisited, but for the present, there are two controversies about the dual visual system that deserve comment.

The first controversy surrounds the exact anatomical location of the purported streams, and the exact functional attributes that can be mapped onto each location. In the introduction I alluded to the fact that I was not interested in the dichotomy proposed by the dual visual systems hypothesis per se, as much as the functions

proposed to be supported by this processing. Theorists such as Glover (2004) and Jeannerod (2006) propose reconceptualisations of the roles of dorsal and ventral – and in some cases, 'ventro-dorsal' (Gallese, 2007) – regions. However, neither the exact architecture of the visual system, nor the precise allocation of functions to brain regions really changes my account. It is enough for my purposes if all agree that the clusters of functions first attributed to ventral and dorsal streams – rightly or wrongly so – do exist and are significant. Even should the neural sources of this insight turn out to be misplaced, nobody in vision science would dispute the importance of processing visual information to guide action as well as to carve a visual world of objects. Those ideas, not any exact neural geography, are what ultimately ground this account.

The second issue is the question of perceptual consciousness. As seen above, it is possible to demarcate the functions of the ventral and dorsal streams without taking a stance on phenomenological visual consciousness, but to ignore the issue in the context of an essay setting conditions on the achievement of self-consciousness would be suspicious, to say the least. The predicament is this. Milner and Goodale, the main proponents of the modern duplex vision hypothesis, have maintained that ventral visual processing, but not dorsal processing, is a necessary condition for consciously experiencing visual scenes. Their famous visual agnosic patient, D.F., was stripped of conscious access to a visual world of objects. Milner and Goodale attributed D.F.'s remaining degraded visual experience to spared areas along her V1-to-inferotemporal-lobe pathway (Goodale & Milner, 2004). Dorsal stream processing was excluded from the story of visual consciousness, not so much because D.F.'s remaining perceptual awareness seemed somehow uncharacteristic of dorsally processed information, but more because the researchers' stipulated role for dorsal visual information made it an inappropriate candidate for entry into consciousness (Goodale, 2005). This reasoning led them to assert that "conscious visual experience of the world is a product of the ventral not the dorsal stream" (2004, p. 109).

What Milner and Goodale have on their side is patients' reports about what they can see. Ventral processing gives rise to reports of seeing objects. Reportability, verbal or demonstrative, has been equated with visual consciousness both by researchers who subscribe to the idea of ventral processing producing visual consciousness and those who do not (Goodale & Milner, 2004; McIntosh *et al*, 2003; Johnson & Haggard, 2005). This equation has been challenged. There may be

conscious but unreportable experiential contents (Nudds, 2007), and reports may reflect analytic knowledge about oneself not directly available in experience (Clark, 2007), as when I successfully reproduce a deviation in my hand's reach though I have no real-time experience of the reasons for, means towards, end state of, or metrics pertaining to the deviation as I act (Johnson & Haggard, 2005). Two kinds of dorsal visual content of the experienced but unreportable variety may be what has been termed *directive content* by Nudds, and the related idea of *presence*. These putative conscious contents contribute respectively to the experience of acting under visual guidance and to the feeling of being spatially enveloped by a scene (Nudds, 2007; Loomis, 1992; Matthen, 2005). Among those researchers who meet Milner and Goodale's reportability trump-card head on are Johnson and Haggard (2005), whose subjects non-verbally reported on an unnoticed reaching detour by reproducing it even as they failed to verbally report the change of course. Without taking an explicit stance on reportability, Gallese (2007) has claimed on separate grounds that dorsal processing contributes to perceptual awareness of space and to action understanding.

Whether consciousness of a visual scene is enabled by the ventral or dorsal stream or both is beyond the scope of this paper. It falls to me, rather, to show that I may safely put this issue aside while still hoping to give a right account of the kind of self-consciousness that develops in a separated, engaged creature. I believe visual consciousness and self-consciousness are orthogonal enough to make this possible.

The contents of visual consciousness may be thought of as renderings of the information implicit in a retinal image that come into contact with knowledge, (motor) know-how, or expectations, perhaps including efference copy information, in order to produce intelligent or appropriate response in the world. Notice that this wording is neutral about which functional areas in the brain mediate visual awareness (cf. the similar but non-neutral wording in Clark, 2007). The contents of the varieties of self-consciousness I will discuss will play a role alongside other knowledge in bestowing survival advantage upon the conscious creature, just like visual experiences do. The difference is that self-conscious contents at least: *a)* implicitly or nonconceptually contain some 'self' content, whether this be a perspectival, agentive, egocentric, or otherwise self-specifying bit of the experience, and at most: *b)* have the capacity to contain, over and above the content in *a)*, an explicit or conceptual or self-referring content. All this remains elusive at this stage. What is central, however, is that visual consciousness is about visual surroundings, whereas self-consciousness is

directed at the experiencing self. Of course, there will turn out to be a link between visually processed information and selfhood. That is the topic of this paper. But the kinds of self-consciousness pursued via these arguments will not lean on the *experiential character* of visual experience at all.

3: Perception for Separation and Engagement

One group of tasks Milner and Goodale administered to visual agnosic patient D.F. involved *pantomiming* actions. Pantomimed actions are delayed actions towards objects which have been removed from the visual scene for long enough that dorsally extracted information, being evanescent and constantly revised, will have decayed. These actions are therefore said to be done *off-line* (Westwood, Chapman, & Roy, 2000). D.F. and other ventrally lesioned persons are good at *on-line* visually guided action, but perform poorly when asked to pantomime. They make to grasp just-seen objects by means of inaccurate grip postures, for instance. Optic ataxics, however, display improved performance when asked to pantomime. For these patients, ventrally encoded information is more helpful than the disrupted information from the dorsal stream that usually commands their actions, and when dorsal information has decayed, they can mobilise ventral information in producing an educated, if imperfect, pantomimed grasp (Goodale & Milner, 2004).

These results cannot be explained solely on the basis that ventral stream information decays slowly enough to be recruited to control ataxics' actions. Rather, ventral stream vision feeds into perceptual memory. When I turn off a lamp and walk away, I can still roughly remember its dimensions and position, and am able to indicate the lampshade's height and girth based on memory. The philosophically remarkable thing about pantomiming actions on the basis of ventral vision is that it nicely demonstrates that objects seen during direct sensory coupling with the world can be unplugged and carried around in the head. What D.F. cannot do, then – what a dorsal stream cannot do – is take visual perception off-line. It is a ravenous and amnesic system; without constant contact with the world that serves it as input, it is crippled. Of course, it is advantageous that dorsal processing be frequently refreshed. That is what makes dorsal information so precious in the here-and-now. Yet the fact remains: whereas the ventral stream can sift away the sensory inundation of the world and collect its favoured nuggets of information, packaging them for pick-up by

memory systems, the information extracted by the dorsal stream cannot survive isolated from an embodied, embedded context.

All living creatures can *engage* with their local surroundings by acting, being acted upon, and generally responding to the world insofar as it is immediately available, examinable, malleable though refractory, and significant for continued survival. Situated roboticists have attempted to instil such engagement skills into their creations, with encouraging results (Brooks, 1991). Many creatures also make contact with the world via the skill of *separation*. Separation is characterised by thought about objects, paradigmatically when those objects are temporally or spatially removed from the thinking creature. In artificial intelligence, separation has points of contact with symbol-crunching von Neumann machines like personal computers. Once situated cognition joined Cartesian dualism among the ranks of valuable meta-theoretic frameworks for the study of mind, analogues of the ideas of both separation and engagement became common themes in cognitive science. To form a rather sloppy (and at times arguably orthogonal) list of similar distinctions, there is: mentality and physicality, representation-heavy and representation-free tasks, thought and action, and descriptive and demonstrative reference.

A thinker without engagement skills would of necessity inhabit a mental realm wholly disconnected from any active existence in a material world. And a creature without separation skills would be stuck in a closed loop with its environment. A main premise of this paper is that a creature with no hope of effecting separation from its environment would fall short of conditions for sophisticated self-consciousness. So far, the borderlands between separation and engagement have not been adequately delineated. I will eventually position such themes as causal effectiveness, representation, ecological perception, visual tracking, and feature-placing in the picture. For now, it is enough to note that occurrences of engagement, broadly construed, are curiously widespread.

According to Smith (1996), the phenomenon of *deixis* is commonplace. *Deixis* is another word for *indexicality*, that is, orientation relative to some origin (Pylyshyn, 2000). It is called egocentricity whenever there is an ego of which to speak. Every physical event, Smith believes, consists of clashes between deictic forces. Magnetic forces pull, towards their centre, iron filings which resist being pulled away from their own centres of gravity. Whether or not the physical substrate is steeped in deixis, it is relatively incontrovertible that physical changes brought about by ordinary material

objects, biological thinkers included, radiate from roughly centred locations. It is easy to hop ten centimetres, and harder to do a long-jump. Note that the skill of engagement requires lots of proverbial hops, whereas separation takes long-jumps. Smith (2001) claims that indexicality lies at the heart of the skill of engagement. And insofar as deixis is 'cheap' in the above sense, while non-indexical activity is expensive, engagement is cheapened compared to separation.[3] Glibly stated, engagement-as-deixis is a plentiful and low-cost; separation is rarer and costlier.

Indeed, it will emerge that separation is not only rare and expensive, but precious. In Brian Smith's words, "what it is to think [..] is to represent the world out there, beyond the periphery, by rearranging your internal configuration, and adopting appropriate habits and practices, so as to behave appropriately with respect to—so as to develop hypotheses concerning, so as to stand in appropriate normative relation to—that to which you are not, at the moment, physically coupled" (2001, ms. p. 10, cf. p. 17). Smith is referring to a very rich notion of separation here. The perceptual memory of a lampshade pales in comparison. Yet there is a strong connection between rich separation and memorable lampshades. In order to draw the connection, more has to be said about perceiving the world.

Some theorists reserve the word *perception* for the processes and products of ventral stream activity (Goodale & Milner, 2004). Others, notably J.J. Gibson (1979), have felt comfortable applying the idea of perception to dorsal-style activity. I believe privileging ventral processing in this regard depends on two theoretical confounds, between the ideas of perception and conscious sensation on the one hand, and conscious vision and ventral processing on the other. Similarly, Gibsonian theorists take a reductive tack, equating perception with direct, pre-conscious loops of acting and reacting in the world. Since neither of these equations should be adopted without justification, I will use *perception* as a blanket term for both kinds of visual function. To make a stronger point, I will take it that visual perception occurs through the general cooperation of ventral and dorsal streams. Recalling our early maxim: the world is one, and metaphysically, perceptual reality is unified. In fact, so too is our phenomenology of acting and observing. If the ventral and dorsal stream were not cooperative, we would sense tension between action and observation. Only in

---

[3] Of course, skilled survival-enhancing engagement is expensive – think of how hard it has been to construct a robot with a natural gait. It is simply the ability to make a mess within a centred radius that is easy. More will be said about 'rich' engagement later.

pathological cases or empirical settings does this occur. Dorsal and ventral streams act in exquisite synchrony to create a perceptual habitat for sighted, embodied creatures.

Despite that conciliatory note, it is time to divide duplex visual perception into its elements, knowing full well that there will ultimately be a consequential reunion party. Given what was noted about the skills of engagement and separation, what can be said about the main contributions of ventral versus dorsal processes? I will contend that ventral processing is *effortless proto-separation*, and dorsal processing is *engagement*. Perception is both engagement and an effortless proto-separation, and is subserved by the two visual systems. The core paradigm for explaining engagement is Gibsonian ecological perception. The core idea behind proto-separation is Jacob and Jeannerod's idea of selection, of the binding of visual properties at locations into what I will hereafter call *proto-objects*, thus downgrading from my previous references to ventrally constructed objects. Proto-objects are "clusters of proximal features that serve as precursors in the detection of real physical objects. What [different uses of *proto-object*] have in common is that they refer to something more than a localized property or 'feature' and less than a recognized 3D distal object" (Pylyshyn, 2001, p. 144n). Let us tackle the ideas of effortless proto-separation and engagement one at a time.

3.1: Ventral Processing and Proto-Separation

It will be worthwhile to sketch some of the properties of ventrally mediated effortless proto-separation. Here, let me introduce Vee. Vee is a hypothetical creature whose non-mammalian brain – including her midbrain visual system – has just evolved a skeletal V1-to-ventral pathway. The key constraint on Vee is that she has no dorsal stream. I will use Vee's case to flesh out the achievements and shortcomings of ventral stream contributions to vision and cognition.

First, in what ways is ventral processing separation? As Smith (1999) notes, attempts to model perception by A.I. researchers have revealed that the world is metaphysically chaotic, "an extraordinarily rich, four-dimensional world (time & space), that instantiates a bewildering array of features – colours, smells, textures, fogginess, whatever – without any concomitant commitment to individuals or identity" (ms. p. 10). To make this messy world tidier, Vee must detach from it ever so slightly, forsaking completely immersed activity. By giving Vee landscapes of

proto-objects (*popjects* for short)[4] perched upon and behind and occluding more popjects, the ventral stream produces the barest state of observer-side estrangement or abstraction away from a scene. In a sense, seeing popjects is taking a step back from reality to get a better view.

These metaphors can be given some grounding. Proto-separation has a standard recipe for the enterprise of popject-making. Ventral vision makes objects out of *visual features* – properties like lines, textural grain, shading, and edges – likely by *binding* or tagging them to locations (Smith, 1999; Campbell, 2006). I quoted Pylyshyn (2001) as saying that popjects were thicker notions than visual features. According to Austen Clark (2004), popjects are moreover thicker notions than *located* visual features, which result from the procedure of *feature-placing*, a coordination of spatiotemporal positions with sensory feature maps. Crafting popjects requires more prowess than feature-placing because the former would overwhelm the modest computational resources involved in attending to or indexing bits of a sensory tableau. Think of how many overlapping or proximal visual features make up a bottle of water. Humans are said to be able to attend to about four items in a visual scene at a time (Clark, 2004). Clearly, carving a real world into popjects involves binding more than four features to each location. This leads Clark to claim that seeing popjects requires more functionality, and further dedicated mechanism, than that which enables feature-placing. I believe this selection mechanism, whether it is attention-like, or memory-like, or a matter of taking the temporally synchronised firing of feature-placing neurons to indicate the presence of a popject, is the paramount and constitutive achievement of ventral visual processing.

In a roundabout way, this helps explain how popjects give separation a foothold. Smith (1996) identifies four ways in which the ability to *non-effectively track*, or attend to what is no longer coupled to a creature's sensory system, foreshadows full-fledged separation. I believe these criteria already apply to Vee's creation of popjects, which precedes non-effective tracking because it does not conceptually necessitate access to memory (we will soon see that practically, popjects are ineffective without the aid of memory systems). Smith's (1996, pp. 222-224) signs of fledgling separation are:

---

[4] *Popject* is simply a play on a shortened form of proto-object, p-object, which highlights the 3D pop-up metaphor implicit in discussions of figure-ground segregation.

1. The newfound coherence of distinguishing between what Vee is doing (in Smith's case, non-effectively tracking; in ours, seeing popjects) and how she does it.

2. The potential for error, which is symptomatic of loosened coupling between Vee and her environment.

3. That Vee must suddenly shoulder commitment to and responsibility for her rendering of the world.

4. That Vee now has use for wiles and mechanisms that extend, improve, or support her rendering of the world.

Wherever ventral selection mechanisms are mobilised to create popjects, it makes sense to discriminate between what Vee does and how she manages it. Vee's ventral mechanisms could have been engineered differently by evolution. And there are different ways for the very same ventral mechanisms to popjectify – different ways in which collections of features can become popjects. There is choice in the matter. What extent of spatial overlap between features is needed? Which features demarcate boundaries and which indicate lighting? Contrastively, it is a crucial point that *visually tracking* – otherwise known as *effectively tracking* (Smith, 1996) – what is continuously in view is such a physically (Pylyshyn, 2000) and conceptually (Smith, 1996; Bermúdez, 1995) foundational ability that there are no choices involved. There is only one way to visually track, by sequentially accessing spatiotemporal points without reference to particular located features (Pylyshyn, 2000). One tracks by tracking locations, and locations have no properties other than being locations. More, tracking mechanisms could not have been engineered differently without changing what the creature could be said to be doing. Tracking by indexing features or attending to popjects or objects would not open a how/what rift for tracking. It would only change the parameters of the act so that it no longer qualified as tracking at all. A similar lack of wiggle-room plagues feature-placing; to bind a feature to a place just is to bind the feature at a point to that point. The creation of popjects marks the beginning of a coherent distinction between what is accomplished and how.

The second criterion, the potential for error, arises because making popjects is a way of representing the world, as I will contend below. As to the third point, Vee must take responsibility for her choice of popject-landscape in a very primal sense.

Making popjects out of the wrong collections of features could get Vee eaten or lead to her starvation. Finally, Vee certainly does have use for extensions to her ventral mechanisms. That is why, when the opportunity presents itself for Vee's proto-separation to become full-fledged separation, evolution favours that change. It will emerge that a fully separated creature just is a proto-separated creature with added functionality emerging from its immersion in a social and encompassing material world. In embodying Smith's four signs, Vee firmly has her feet on the road to separation.

How far from that destination does Vee stand? That is, what makes her skill one of *proto*-separation? Viewing scenes as composed of popjects is not *full-fledged separation* because popjects cannot easily defy their present spatial contexts. A popject cannot defy its visual context because, as Smith's (1996) says, it is not properly "'sedimented' or 'extruded'," by the perceiver, "as a discrete autonomous individual [..] and thereby locate[d] in the wider world" (p. 225). It is as though proto-separation crafts engagement-based objects, but it is these same popjects which will become the thick objects that populate our thoughts.

The close coupling of popjects to a present locale may seem at odds with my discussion of perceptual memory. Yet an optic ataxic's ventral stream setup will have points of departure from Vee's. Whether a skeletal ventral stream would have access to, or be accessible to, perceptual memory is a matter of speculation. Nonetheless, to illustrate perceptual popjects' lack of 'sedimentation' – their instability – we may consider two possibilities. First, imagine Vee's popjects cannot enter into perceptual memory. Without memory we could not visualise popjects without direct sensory access to them. Vee's popjects would disappear every time she shut her eyes. Clearly, these would not be discrete autonomous individuals. More telling is the case where Vee's popjects are granted access to perceptual memory. In this case, popjects would *still* be unstable, because there is much more to objecthood than simple formal coherence. For one thing, Vee would have to conceive of her popjects as causally efficacious entities. Much more will be said about the difference between popjects and objects when the conditions of full-fledged separation are explored.

Finally, what makes effortless proto-separation *effortless*? It is the bare function of ventral processing to present Vee with a visual world of popjects. No further reasoning or decision-making is needed for this to happen. This is implicit in the neural architecture. Maps of visual features sent from V1 are integrated right

inside the ventral stream (Goodale & Milner, 2004). An example may be illustrative. D.F.'s brain disorder is specifically *apperceptive agnosia*. She can see visual features with great acuity, but her brain does not build objects out of these. In contrast, *associative agnosics* have a higher-level disorder. They see objects, but cannot categorise or name them (Nudds, 2007). They do not have problems with the more fundamental skill of effortless proto-separation, being adequate at drawing objects or matching them. There are no apperceptive agnosics who nevertheless have associative abilities because labelling occurs downstream of ventral processing. Naming and categorisation necessitate cooperation between the ventral stream and higher-level areas of the cortical brain. They are effortful inasmuch as they require extra processing of visually pre-processed information.

　　As well as resulting from perceptual selection and being allocentric and relational, ventral stream content is *presentational*. Presentational content represents the way things really are in the world, independent of observer attitudes (Nudds, 2007). Notice that once full-fledged separation is achieved, once objects can be unplugged from environments and thought of as autonomous individuals, they lose this claim to veridicality; there can no longer be said to be an overarching *mind-to-world* direction of fit between those contemplated objects and reality (Jeannerod & Jacob, 2005). As a fully separated creature, I can imagine objects that do not exist, and I can imagine real objects existing in hypothetical arrangements.

　　Proto-separated content is also *representational*. It may not create the richest of representations, ones "whose functional role is to act as *de-coupleable surrogates* for specifiable (usually extra-neural) states of affairs" (Clark & Grush, 1999, p. 8), but it results from an almost paradigmatic act of information *dilation* and *compression*, whereby raw data is given more or less consideration by a processing system. Ventral processing dilates information about localised features collected at a location while compressing less important noise like the precise wavelengths hitting the retinal surface. For Clark and Toribio (1994), information dilation and compression are marks of a system that is at least modestly representational. Crucially, dilation and compression allow proto-separated representations to be erroneous, therefore fulfilling Smith's second criterion. Ventral vision may be presentational, but at a level of description below that of observer attitudes, it does take liberties in parsing a scene. This flexibility is one of the culprits behind visual illusions. Lastly, representations created by Vee's ventral stream are *nonconceptual*. These are representations that do

not make direct use of *concepts* – those mental representations that are flexibly recombinable constituents of thought (Kirsh, 1991).

I mentioned that proto-separation creates an informational overlay and therefore a visual evidential base to which decision-making processes can look in producing visually guided judgements. But what about the very simplest of visual judgements, those of recognising and reidentifying objects? Are those handled inside Vee's ventral stream, or do they require inferential or rational capacities?

The ideas of recognition and reidentification lack universal interpretations. I will adopt the term *recognition* for repeated visual identification which does not require thinking of the recognised object *as* any particular thing. Especially forbidden is recognising the object as something exhibiting *object permanence* – persistence in the world across a break in observation (Meltzoff & Moore, 1995). In fact, there is one description that may be permitted in the case of recognised objects, that of being 'seen before' or 'familiar' or 'the same'. Strawson (1959) appears to use the word *recognition* in this way (p. 76). *Reidentification*, on the other hand, I will take to entail the observer's positing a world out there such that a concrete object, twice encountered, must have been at sequential spatiotemporal positions in that world throughout its absence from view. As Dennett strikingly puts it: "Many organisms 'experience' the sun, and even guide their lives by its passage. [..] But we human beings don't just track the sun, we make an ontological discovery about the sun: it's the *sun*! The very same sun every day" (1996, p. 58). For Strawson, recognition is a perverse breed of identification, since repeated identification and knowledge of a continuous external reality are to his mind tightly connected. However, I will illustrate that mere recognition is the only kind of reidentification applicable to popjects.

Vee will have recognition capacities due to her ventral stream that a more primitive creature – a purely midbrain-sighted one, say – would not. Here we must make a two-pronged amendment to the story, gifting Vee with a memory and attentional system, and admitting memory and attention into the category of cognitive functions necessary for full-fledged separation (Kirsh, 1991; Campbell, 1995). Attention is necessary for whatever ventral mechanisms produce popjects. Memory is necessary for recognition, and the importance of recognition should come as no surprise considering it is a precursor to reidentifying individuals, a skill given pride of place in numerous accounts of cognitive development (Strawson, 1959; Evans, 1982;

Dennett, 1996). With memory and attention abilities, Vee should to least be able to recognise popjects whose appearances are stable (perhaps because Vee's body remains stationary) over short spans of time. Indeed, in humans and monkeys, there is evidence that certain ventral stream neurons are significantly more active for recognised faces than detected but unrecognised ones (Grill-Spector, 2003). Even supposing a large change in perspective on a visual scene, or a mutating popject, Vee should be able to register the familiarity of popjects because of the relational positions they occupy in allocentrically processed scenes.

What Vee cannot be assumed to have is the ability to reidentify. She cannot yet be assumed to have a conception of external reality and stable objects moving and existing within it. According to Strawson (1959), reidentification, object permanence, and external reality are jointly earned. "For on the one hand places are defined only by the relations of things; and, on the other, one of the requirements for the identity of a material thing is that its existence, as well as being continuous in time, should be continuous in space" (p. 37). Vee's world is more sophisticated than that of a frog tracking its moving prey, but less so than that of a four year old human. Perhaps she totters between recognition and reidentification like a human infant:

> Before objects are permanent, infants conceive of the external world as what is present now. Nothing is hidden because there is nowhere else to be. However, this does not mean that they are solely influenced by the present. Representations constructed from previous encounters are maintained in memory, and current perceptions may be interpreted in terms of them [..]. These memories create an internal realm, but before things are permanent, these memories do not refer to an enduring world (Meltzoff & Moore, 1995, p. 60)

In the next section, I will put forward conditions which would enable a proto-separated creature to develop into a fully separated one, able to entertain thoughts about autonomous individual objects in a continuously extant material world.

3.2: Becoming Separate

In preceding sections, I have alternated between two construals of *full-fledged separation*. On the one hand, I claimed that full-fledged separation involves thinking about true autonomous objects, not perceptual popjects. On the other, I emphasised that full-fledged separation is reflection on what is spatially or temporally removed

from the thinker. These descriptions merge when we think of separation as the ability to have conceptual thoughts.

Let us consider some species of thoughts. Imagine a creature having a thought like

Firas is quick,

where quickness can be thought about only whenever Firas is thought about, or,

where quickness cannot be thought about whenever foxes are thought about.

Such a thought would disobey Evans' (1982) Generality Constraint. This famous principle dictates that grasping sophisticated thoughts requires the exercise of two simultaneous skills. To have a thought respecting the Generality Constraint, say, *Bassel is wise*, I must be able to attribute wisdom to a variety of objects, and I must be able to comprehend a spectrum of thoughts about Bassel independently of thoughts about wisdom (Cussins, 1992; Evans, 1982). Adherence of thought to the Generality Constraint is generally considered to be indicative of the presence of conceptual content (Evans, 1982; Gillett, 1987; Smith, 1999). Recall that I take concepts to be nothing more than mental representations or thought elements that may be recombined and flexibly rearranged. Concepts, I am assuming, need not depend on linguistic capacities. Therefore, in adopting singular terms and predicates as the building blocks for general and conceptual thought, I will hold that they may be object and property representations constituted by visual or motor imagery just as they may be linguistic names and properties.

Separation is grounded in a skilled kind of access to the perceptual world. What links such an ability to having general or conceptual thoughts? This question cuts to the heart of what full-fledged separation is. Becoming separate is an acquisition of thought elements, gathered and tended in the perceptual world, which can be flexibly deployed in contemplation, often off-line from perception. Jesse Prinz's neo-empiricist view that "all (human) concepts are copies or combinations of copies of perceptual representations" (2002, p. 108; cf. 2007) is almost right. For singular terms to be independent of property predicating terms, however, the objects they refer to must be comprehended not as unstable popjects and popject features, but as persisting individuals independent of a perceived scene. Popjects are nonetheless essential to this enterprise. It is popjects which must be stabilised and sedimented into objects.

True objects have dimensions far surpassing their spatiotemporal appearances. In referring to Chris, I refer to Chris who is a collection of molecules, Chris who was at home yesterday, and Chris who will be excited about the concert all with the same act of reference (Smith, 1996). Indeed, to extrapolate Frege's (1980) Context Principle, understanding objects just is to understand them in situations, doing things, being certain ways. The recombination of conceptual elements entails that singular terms or predicates can be unplugged from a mutual coupling and grasped to have or lack a range of properties, or to belong to numerous objects, respectively. This correspondence suggests the link between separation and conceptuality. Objects and object arrangements are the recombinable furniture of the world, and the singular terms and predicates representing them form the recombinable furniture of conceptuality.

Not only objects but also their arrangements must be stabilised in conceptual thought. Understanding predicates, say grasping the notion of *quickness*, requires a stabilisation and sedimentation of the different states and various articulations that objects can embody. This crucially involves grasping that objects can be at different spatiotemporal positions in the world, and that they are both efficacious and vulnerable: they both causally affect and are affected by the world. Notice that the capacities to conceptualise objects and their arrangements are independent but vitally intertwined. Grasping what kind of states predicates refer to depends on understanding something about objects (and not just seeing popjects), and grasping what objects are requires imagining them in different places and different times, causally interacting with different things. Indeed, these are the two cognitive skills Evans alludes to in describing the Generality Constraint, except that these skills are more than 'cognitive'. They are skills deriving from immersion in a material world. To have true objects in mind, a creature must share an enriching co-existence with them.

Material objects begin life as proto-separated thinkers' unstable popjects. The case is different for abstract terms and predicates unlikely to be brought before the visual system. Consider the thought *Truth conquers all*. The concepts involved do not begin as popjects. Perhaps the concept *truth* results from stabilising the directive superiority of a thought about objects and states that obtains in the world over graspable but non-presentational thoughts. The concept *all* may be a stabilisation of the total domain of objects available in completing an unsaturated thought, like *x is*

red. To stabilise abstract relationships, general thoughts about true objects must already be in place, and likely linguistic competence too (Cussins, 1992). These conceptual capacities cannot be explained by the more primitive task of extruding objects, then. They are post-separation capacities. Stabilising objects and arrangements is sufficient for conceptual *thinking*, but not for conceptual *mastery*. Thoughts about immaterial or abstract objects may demand full-fledged separation skills in concert with other factors.

Incidentally, this poses no problem. Even as a fully separated creature, I may have trouble conceptualising things or pairing concepts together. Associating a concept sedimented from perception, like *chair*, with an abstract one like *being made of quarks*, to form *The chair is made of quarks*, may be beyond me if I misunderstand what quarks are. My thoughts are still general, though, since *chair*-thoughts and *made-of-quark*-thoughts do not directly preclude each other in my thinking. I grasp what it would be for a chair to have various properties, and once I do learn the concept *made of quarks*, I learn it insofar as I can apply it to a multitude of singular terms. This echoes the case of all abstract concepts, which may come later to a conceptual creature than perceptually salient concepts.

With respect to the two construals of separation, it has been established that being separate *qua* thinking about objects makes for a conceptual thinker. General thought about objects is furthermore essential for reflection on what is spatially or temporally removed from the thinker. It is none other than objects and their arrangements that are thought about off-line from perceptual input. Object and property representations comprise all thoughts about what is distally or temporally remote. Rehearsing nonconceptual perceptual memories would be a specious example of separated, conceptual thought. Just as *quickness* could not be thought about whenever *foxes* were thought about, non-perceptual elements would be precluded whenever perceptual elements were rehearsed, in violation of generality. Additionally, thinking about perceived elements as true objects involves more than ventral vision. When I reason that this water bottle would make a good bludgeon for a nearby pest, I am exercising conceptual separation abilities in the here-and-now, thereby detaching from what is presented to me in vision. Finally, it is objects and their arrangements which will matter enough to a creature that it will direct thoughts at them when they disappear; popjects are too ephemeral to shoulder such influence.

A separated creature thinks about more than it perceives. It thinks about objects and the ways objects can be, and in doing so it must necessarily be thinking of them as general and variable, as more than what they appear to be. And though it thinks about more than it perceives, the creature begins by modestly perceiving simple popjects. How, then, does such a rudimentary kind of thinking as seeing unstable popjects, perhaps even remembering them, blossom into generalised thought about multilayered objects and, what is equivalent, into conceptual, general thought? Here is a survey of some theoretical proposals.

Social and cultural immersion probably have something to do with it. Dennett (1996) reasons that the advent of communication heralds the achievement of conceptuality with its upshot of linguistic thought about 'thicker' objects and properties. Cussins (1992), too, gestures towards language: "One familiar and important way in which stabilization is achieved is by drawing a linguistic blackbox around a feature-space: the imposition of linguistic structure on experiential structure" (p. 679). However, it is unclear to me that syntax or grammar emerge prior to separated thinking rather than because of it, or concurrently with it. Perhaps, then, something like a diffuse social atmosphere where individuals are not always in direct contact with one another is a better prior condition. For Dennett, such an atmosphere makes it possible and advantageous to fabricate reasons for one's behaviour and self-attribute them, thereby triggering the development of language.

Self-attributions of this kind appear to be on the right track, since they just are thickenings of the idea of objects of a very special sort – *persons* (Strawson, 1959). Smith's four early signs of separation support the importance of membership in a society of conspecifics. Social immersion may, for instance, allow for a profound kind of responsibility and commitment to one's role in the world. Commitment to the world would drive a normatively guided thickening of popjects into objects even as attributing new significance to objects fostered new commitments. My role in society prohibits my treating others as mere visual apparitions. I cannot throw darts at a fellow creature the way I might deface a photograph without incurring serious consequences.

Commitment and responsibility can be exported from a social context and distilled into a kind of discriminating world-directedness. Avila-García and Cañamero (2005) describe how simulated hormonal modulation can influence a robot's choice of actions in its environment by directly colouring its sensory input. We may compare

this to the child's game where a player's goal-directed navigation is controlled by an accomplice's shouts of 'hotter!' and 'colder!'. Goal-directed behaviour presupposes this sort of differential relationship with the world. It figures into both the selection of goal states and of means to those ends. Commitment and responsibility, too, are at once consequences of taking, and reasons to take, certain bits of the world to matter more than others. Via this discriminating, world-involving gaze, popjects may gain some of the depth of true objects (Smith, 1996). It is, after all, part of the ontological status of my teddy bear Teddy that he is one of my oldest possessions, that I am fond of his mangy appearance, that I *care* about him.

Mining the physical realm for insight into the achievement of separated thought is just as revealing. According to John Campbell (1995), "[t]o be using a term to refer to a concrete object, one's reasoning using the term must display a grasp of the two dimensions of the causal structure of the thing: being internally causally connected over time and being a common cause of many phenomena" (p. 32). Ideas about objects being causally connected over breaks in observation will be inextricable from a grasp of object permanence and the existence of an independent, persistent world. Whatever new mental prowess allows a creature to reidentify objects as well as recognise popjects will play a role in the genesis of separation.

Vee is at a loss when it comes to securing the above ingredients for separation. A ventral visual system could not possibly underlie all this richness. Yet, a partial solution might not be far removed from ventral vision. Some ingredients that go into full-fledged separation will ultimately be secured by visual processing. We need only turn to the dorsal system.

3.3: Dorsal Processing and Engagement

Our assortment of thickening agents, from understanding causal physical structure, to being able to reidentify objects and knowing that reality is mind-independent, to being differentially world-directed, cannot be explained without invoking the functions of dorsal processing. The significance of societal context, and perhaps even the dawn of linguistic communication, will also implicate vision-for-action in the developmental tale of a separated creature. For the connection between dorsal visual processing and separation to neatly rise to the surface, an elaboration of the properties and constituents of dorsal processing is in order.

Assume that there could be a creature whose midbrain visual system has just evolved a dorsal stream, but not yet a ventral stream. This dorsal counterpart of Vee is named Dai. What is it for Dai to perceive the world? In a 2002 paper, Joel Norman explicitly drew strong parallels between the features of dorsal-style vision and the Gibsonian notion of ecological perception. Such a connection is implicit in several psychological, philosophical, and computational analyses (cf. Butterworth, 1995; Chemero, 2007; Hatfield, 1990; Jordan, 2003). For Gibsonian theorists, Norman contends, "[p]erception is an achievement, a keeping-in-touch, not a passive experiencing of one's conscious responses to stimulation. This view contrasts with the constructivist perspective of a perceiver who passively examines her/his conscious awareness of the stimulation impinging on her/his senses. This view of perception as resulting from an active perceiver is, of course, consonant with what we know about dorsal system functions" (2002, ms. p. 19). It is not only the link between perception and action that ties the ecological approach to dorsal vision, however. Here are a few more commonalities.

Dorsal perception may be thought of as the perception of Gibsonian *affordances* for action specified in the optic array (Norman, 2002). Says Gibson: "The *affordances* of the environment are what it *offers* the animal, what it *provides* or *furnishes*, either for good or for ill" (1979, p. 127). For instance, something might be sit-on-able or climb-on-able or graspable. Affordances are contained within the sensory tableau. "The basic properties of the environment that make an affordance are specified in the structure of ambient light, and hence the affordance itself is specified in ambient light" (Gibson, 1979, p. 143). It makes sense, then, that the egocentric invariant measures processed by the dorsal visual stream are those same affordances which specify the parameters of graspability, or, say, climb-on-ability, for the use of Dai's motor systems.

Direct, ecological perception is deictic and nonconceptual.[5] With self- and world-specifying information doubly inhering within the format of the creature-centred optic flow (Bermúdez, 1995), ecological perception is intrinsically deictic and even egocentric, insofar as the ego may be unreflective and implicit. Ecological perception involves no concepts, being "fundamentally a process of extracting and abstracting invariants from the flowing optical array" (Bermúdez, 1995, p. 162).

---

[5] That is, conceptuality is beyond the scope of both proto-separation and engagement. It is only available to a separated thinker.

These invariants are not recombinable representations of objects and properties, but context-specific measures to be fed into action-producing cortical areas (Norman, 2002). They are abstract only in the sense that they must be transformed from eye-invariant information to limb-invariant measures, possibly via the mediation of a canonical 'master' egocentric invariant (Cohen & Andersen, 2002). These points overlap with features of dorsal processing. Additionally, just like Dai's bare dorsal processing, ecological perception is amnesic; what is directly perceived is bound to the present, taking full advantage of the optic flow by 'resonating to it' without the interference of explicit memories or predictions (Norman, 2002).

It is unavoidable that the dorsal stream make use of feature-placing and effective visual tracking. Though Dai's brain cannot use features to create popjects, Dai directly perceives body-centred patterns of visual features. Sifting through such low-level properties of a vista as features and positions is, Smith claims, "necessary in order to control the fine-grained detail of action and perception. If you want to place your finger just there, if you want to track that pattern of animal motion against a background of grasses waving in the wind, [..] then the rich detail provides critical resolution for fine-grained action" (1999, ms. pp. 10-11). Feature-placing is perfectly fitting for ecological perception. Located features are resolved into affordances by dorsal vision. This does not involve binding features into spatiotemporal popjects. Rather, feature patterns that make up affordances may be spatially diffuse, cross-cutting objects in the scene.

Ecological perception is a powerful framework for understanding Dai's perception of the world. However, properties have been attributed to dorsal processing which remain unelaborated within the ecological approach. Jacob and Jeannerod (2005) claim that dorsal perception is neither helpfully conceived of as conceptual nor nonconceptual but as *visuomotor*: geared for visually guided action, egocentric, absolute in frame, and operating efficiently on a very limited range of features extracted from the scene (Bermúdez, 2007). This echoes Nudds' (2007) recommendation that dorsal content be considered directive. Several recent accounts of the dorsal stream have attributed a great deal of cognitive complexity to its functions. Jeannerod (2006) argues that dorsal processing is part of a system for motor cognition, and is involved in the production of action representations, motor imagery, imitation, and tool use. In fact, Jeannerod wishes to revamp dominant depictions of the dorsal stream, replacing them with the idea of a system which distils

conceptual sensory and proprioceptive representations into nonconceptual action specifying and controlling measures. Gallese (2007), too, implicates dorsal structures in spatial awareness and action representation.

Is Dai's visual processing as sophisticated as Jeannerod and Gallese would like, or have separation-based skills leaked into the discussion of what is supposed to be active perception in a present and body-centred environment? It is hard to say where the boundaries of engagement lie. For we might think it overkill to marshal dorsal processing to explain simplistic causal interplay with an environment, where survival is irrelevant. A midbrain visual system ought to be more than sufficient for that (Eilan, 1995). Even a weathervane or a Watt governor would be proficient at that kind of engagement! Yet Jeannerod's roster of dorsal skills, including conceptual action representations, seems too substantial to count even as survival-oriented engagement. Here the pressure to re-admit dorsal processing into the story of full-fledged separation is supplemented by an equal pressure to restore ventral processing to its role in evolving and enriching bare engagement into rich engagement as we humans know it. We need to upgrade from weathervanes to ballerinas. After all, dorsal and ventral vision historically evolved in tandem (Goodale & Milner, 2004). We must therefore take leave of our metaphysically unlikely specimens, Vee and Dai, and reunite the functions of dorsal and ventral processing, and the skills of separation and engagement.

4: Interlaced Contributions

What is the role of dorsal processing in achieving full-fledged separation? Recall the factors thought to contribute to turning popjects into conceptual representations of objects in a permanent world:

1.   A grasp of concrete objects as intrinsically causally coherent over time and as common causes for events. A grasp of object permanence and world permanence. The ability to reidentify as well as recognise objects.

2.   Differential world-directedness, including goal-directed action, responsibility, and commitment.

3.   Social immersion: language, or an environment of conspecifics.

Before we treat these specific recommendations, a more general point. If Gallese and Lakoff (2005) are right, many of our concepts, even abstract ones, are *embodied*. They result from activating sensorimotor parameters for action in dorsal cortical areas. This resembles Jeannerod's (2006) contention that all representations of actions activate visuomotor areas of the brain. We certainly cannot have a full conceptual suite without conceptualising the states objects can be in and, if they are sophisticated enough, the actions they can perform. If these researchers are right, and there is empirical evidence that at least concepts for action and intention are embodied (Jeannerod, 2006), then detached proto-separated representations cannot secure all concepts. We need skills of engagement to conceptualise objects moving, doing, being affected, being arranged. This is a first endorsement of the role of dorsal processing in separation.

Whereas popjects are collections of located visual features, concrete objects are causal and temporally extended, with causally connected parts. Campbell (1995) makes this a condition on our acts of reference: "We do not think of a concrete object as simply a collection of features. This shows up in the fact that for one to be using a singular term to refer to an object, there must be a certain density and structure in the conceptual role one assigns to the term" (p. 31). Infants come to think of themselves – their bodies and mental states – as coherent and stable common causes, Campbell continues, by imitating others while experiencing an awareness of the acts of their own bodies. One might wonder whether Strawson (1959) was right to claim that reidentifying persons qua mental and physical is conditional on first reidentifying material bodies. Meltzoff and Moore (1995) maintain that babies' knowledge about human bodies and actions, and budding knowledge of objecthood, enrich each other. Imitation, being present from birth, fosters an understanding of self and other as concrete objects. As Meltzoff and Moore's discussion implies, Strawson's order of events may be too restrictive. Perhaps reidentifying person-bodies at the outset, via a visuomotor imitation-based perception of them, yields hypotheses about the permanence of the world and all its concrete objects.

Due to two-way imitation, infants discover that human bodies are a) internally stable over time – the same tongue protrusion emanating from the recognised popject on two occasions is good evidence that both popject and baby have remained internally stable insofar as the baby has the same phenomenology when she and the

popject produce the expression, and b) common causes of events – when the baby moves her tongue and opens her mouth, she can sense that those actions belong to the same imitative exchange with another human body and require the same effort. In fact, Campbell (1995) says that knowledge of causal structure is tied up with, and may induce, some sense of object and world permanence. Causal coherence requires that experiences of my body vary in systematic, non-radical ways across time and across changes in my spatial locations, which may lead me to conclude that I continue to exist when I fall asleep, for instance. And if I am intrinsically similar to the visual popjects I imitate, I may extrapolate that interpretation to the world at large. Dorsal vision for imitation, in these ways, may help turn visual popjects into causal, permanent objects.

What about caring about the world, or having a compelling differential perception of it? I believe that dorsal stream processing contains within its brand of perceptual processing the very spark that makes a creature respond differentially to its surroundings. Vision-for-action just is vision that constitutes a driving force for the perceiving creature. It is the accomplice screaming 'hotter! now, colder!'. Dorsal vision is directive, computing a landscape of competing affordances, some pointing the creature towards benefit, some directing it away from disadvantage, and all enabling goal-directed behaviour. And because the dorsal system is visuomotor, with heavy interconnections with motor cortex (Jeannerod, 2006), this world of affordances suffices to compel a creature to avoid a predator or climb a berry tree – in short, to *care about itself and its world*. This secures a ground for commitment to and responsibility for one's own survival. As for commitment and responsibility towards others, what, if not the ability to have inner sensations correlated to what is seen, in the first instance because of imitative capacities and mirror neuron function, could subserve such significance? Several philosophers, famously Perry (1979), have also noted that indexicality and perspectivalness are motivating forces for action. It is not until I consider that the sinking ship *Atlantis* is this keeling ship beneath my feet that I am compelled to find a lifeboat (Pylyshyn, 2000). Egocentric processing of visual scenes is dorsally mediated, a feature of engagement, not separation.

Turning to social factors, there is evidence that any major implication of social immersion must begin with visually guided action in the form of imitation (Meltzoff & Moore, 1995). Imitation appears to be an innate mechanism for figuratively questioning the world, beginning with caretakers. Meltzoff and Moore argue that

imitating expressions is a discrimination tool for infants. By feeling their own faces automatically contort in the way daddy's smile does as opposed to mommy's pout, babies take the first step to populating their worlds with conspecifics and reidentifying persons as permanent objects, sometimes held to initiate the achievement of a permanent reality (Strawson, 1959). Imitation involves medial brain regions but also dorsal areas (Jeannerod, 2006). Word semantics, too, has been claimed to have dorsal roots. "[P]erception of a word," says Jeannerod, "would automatically activate the mechanism for recognizing and identifying the corresponding reality: action words would activate the motor system to the same extent as visual words would activate the visual system" (2006, p. 163). This might mean that linguistic reference simply cannot emerge without visuomotor processing.

What this discussion demonstrates is that, one way or another, separation relies on dorsal processing and the engagement skills it enables. Does ventral stream processing have a similar role in engagement? I mentioned before that dorsal processing seems too powerful to be pressed into the service of mere action and reaction in an environment. Pure engagement need not be anything other than deictic causal interplay with an environment; at most it is causal interplay with the goal of survival. A midbrain visual system could provide that functionality. This sort of purely engaged creature would not end up with capacities for action representation, imitation, embodied concepts, motor imagery, and the like. Purely *dorsal* creatures would not end up with these rich capabilities. It is a sign of the cooperation of ventral and dorsal processing that dorsal vision is as rich as it is. The engagement we experience, which Dai could not experience, is rich, ventrally augmented engagement.

Popjectification is conceptually prior to rich engagement. Let us suppose pure engagement could call upon survival-oriented visual tracking and feature-placing, but not popject-making. This seems reasonable, as we have found that both Vee and Dai should have these primitive abilities due to midbrain processing or computation in V1, probably with contributions from low-level attention or memory systems. Armed with only these capacities, it would be impossible for a purely engaged creature to have action representations or experience motor imagery. Representations of actions, as opposed to trackable movement, must presuppose at least rudimentary ideas of intentions; and motor images, like embodied concepts, cannot be free-floating arrangements of movement impulses (cf. Brewer, 1995). Motor activity and action are what *things* achieve. While *things* need not in this case be resolved into thick objects,

at least a popject canvas is presupposed by Jeannerod's conception of action-mediating systems as cognitive. If Smith (1996) is right, causality operates at the level of bare-bones deictic flux, and not objects. But action is not mere causality. One surely cannot understand motor control or goal-directed movement without first having a skeleton of coherent amalgams of features on which to culture richer properties like intentions. Just as location serves in enabling feature-placing, the ventral stream will perform the role of binding dorsally involved cognitive representations of agency, motor function, intentionality, and so on, to things that happen to possess them.

Fledgling engagement is soon enough enriched by fledgling proto-separation. Other tools for thinking, especially language and knowledge-sharing in a wide cultural context, may also contribute. The end result is a creature able to engage its world quickly, efficiently, but with internalised benefits from seeing the world as carved into objects. The same creature displays true separation, including conceptualisations of action, of causally thick objects, and a compelling motivation to selectively represent and alter its world, while taking responsibility for those choices. This creature cares enough to think about what is not locally and presently available in a sensory array. And while raw engagement might be uninteresting, both rich engagement and full-fledged separation are vital, interlaced achievements. Such talents, unsurprisingly, make contact with issues about self-consciousness. I will turn to these questions now.

5: Separation, Engagement, and the Achievement of Self-Consciousness

As a point of contrast, I will begin with a sketch of the kind of self-consciousness that might arise from dorsal contributions associated with survival-oriented engagement. This is not a frustrating resuscitation of Dai as much as a question worth asking about non-mammalian creatures. Ecological perception has been thought to afford creatures a humble type of self-consciousness (Gibson, 1979; Zahavi, 2002). Perceivers like these position patterns of visual features in the world and react sensitively to them. All this action and perception emanates from an inherently perspectival perch. By seeing affordances, these creatures demonstrate a *causally indexical* comprehension of their surroundings, appreciating the relevance of

what they perceive for their own actions (Campbell, 1993; Bermúdez, 1995). Perspective smuggles a vestige of self-consciousness into experience. For some theorists this self is mundane: the visual occlusion produced by the shape of my nose allows me to perceive my bodily self within my every perception of the world (Gibson, 1979). It roughly maps onto what Legrand (2007) calls *the transparent self*, what Eilan (1995) conceives of as a raw point of view, and what, for Bermúdez (1995) is a primitive subject and an egocentric origin.

This self is conscious of the world from a unique perspective, but cannot turn its periscope onto its own person – onto itself as mind or body. This is because the creature cannot yet think about objects, and lacking conceptuality, it cannot think of itself as an autonomous, discrete individual among others.[6] "If there are no experiences with contents of objects and places," notes Cussins, "then there are no resources for distinguishing between the contribution of the world and the contribution of the subject" (1992, p. 669). Lacking detachment skills, the creature fails to pull itself free of a self-world coupling. Strawson (1959) would call this creature *solipsistic*; it gets by without separating out its own contributions to perception from those of the world. While this consciousness of the world subsumes a perception of the self, and is bounded by the perspectival nature of spatiotemporal position and the dimensions of the body, it is not a consciousness which may reflect upon the contributions to experience which stem from that perspectival self. Nevertheless, this selfhood is by no means insignificant. We have already seen that indexicality helps secure separation. Moreover, without the self-world perception of the primitive engaged subject, there would be no bedrock subjective mentality to sediment a popject into a richly self-conscious individual (Bermúdez, 1995).

The engaged self is primitive, but engagement skills, having been appropriated into full-fledged separation, additionally enable the *substantive self-consciousness* of a separated creature to arise. Substantive self-consciousness is an awareness of the self as "a persisting object, which is picked out when we refer to ourselves using 'I'. Self-consciousness is a matter of representing oneself as an object" (Eilan, Marcel, & Bermúdez, 1995). As we want to allow for pre-linguistic separation and conceptuality, the language-use characterisation is superfluous. This is a minor alteration. What is important is that the creature be able to think of itself in a general

---

[6] Though the creature might be able to recognise its limbs, it will not be able to reidentify them.

way, directing thoughts onto itself as an object. In utilising the concept of itself as a person, physical and mental at once, this creature is finally *non-solipsistic*, distinguishing itself from the world in thought (Strawson, 1959).

How would separation skills result in framing oneself as an object among others? Our evolutionary predecessors probably took the ventral visual route, beginning by popjectifying their own visible bodies. Actions, intentions, and causal indexicality would have been bound to this popject via a mix of visuomotor and ventral perception. The phenomenology of the perspectival, transparent subject – the self implicit in ecological perception – would then start to become associated with the condensing object. Burgeoning substantive self-consciousness would have been helped further along by imitation capacities, cementing a sense of self as temporally and spatially extended. Imitation could also help fill visual gaps: by popjectifying another human body from an allocentric perspective as I imitate its expressions, I come to know that I am more than a four-limbed torso and a nose-shaped wedge. Realistically, there is more accruable information about the self than vision can afford. Kinaesthesis, proprioception, and feelings of effort would also have participated in this drive towards objecthood and personhood.

Of course, once separation becomes its standard mode of thinking, a creature need not take the long and winding route up from popjectivity to self-objectivity; it need not see its limbs to come to know itself as an object. It may begin with its causal structure as an agent, or with its similarities to an imitated person, or with a sense of commitment and self-preservation, and infer objecthood from those.[7] Or it may begin even higher up, with a more abstracted logical notion of self, as the originator of thoughts which are not perceived to have sources in the external world, for instance (Prinz, 2003). Being conceptual, the creature's grasping any of these stabilisations of self would amount to its knowing itself as object; autonomous, discrete, and capable of possessing many properties and being in many states.

Substantive self-consciousness may fail to capture denser notions of selfhood, however. Augmentation by various cultural tools for enhanced thought, resulting in expanded self-conceptions and (meta)physically expanded selves (Clark & Chalmers,

---

[7] Why should a separated creature have inferential skills? Inference is implicit in the skills honed in on by the Generality Constraint. We may exemplify the generality skill exercised in grasping *a is F* as follows: *a is F* $\rightarrow$ [$\forall x \subset$ D (*x is F*), where D is some appropriate and general domain of objects]. Whenever the conceptual creature grasps the antecedent, it infers the general consequent.

1998), does not directly follow from being a substantive, separated self. With the possible exception of language, these self-creating extensions call on capacities that are not presupposed by separation. Dennett claims that "[w]ords and other mind tools give a Gregorian creature an inner environment that permits it to construct ever more subtle move generators and move testers" (1996, p. 132), where a *Gregorian creature* is one that can refine its thoughts based on designs and structures in the external world (Clark, 2002). The separated creature is not by rights a Gregorian creature. All it can do is stabilise the world by bringing its thoughts and actions to bear on it. The Gregorian creature does something more advanced; it stabilises and streamlines its mind by bringing culture and the environment to bear on it. We can imagine a separated creature who is not yet Gregorian, but a Gregorian creature must first be separated, because it must conceptualise the world as a world, of tools as tools, and of communication as an exchange of ideas between individuals. Likewise, without leaning heavily on enculturation, separation cannot explain the so-called *narrative view of self* nor Wolfgang Prinz's self-morphic state, both of which liken the self to the protagonist of a temporally extended autobiographical tale (Prinz, 2003; Proust, 2003). Self-narratives derive from oral traditions in mind-external cultural contexts.

The substantive self-consciousness of the newly separated creature is the most basic form of reflective self-awareness. Indeed, we find that even barer, nonconceptual, engaged notions of self, *if they are to be reflected upon as well as present experientially*, have to be preceded by separated self-consciousness. Reflections on the self as an ecological self, as a raw point of view, or as a *performative self* or agent (Legrand, 2007; Stern, 1985) are post-separation varieties of self-consciousness. We must grasp personal objecthood before we can grasp, say, egocentricity, which is a property of persons. This harkens back to the argument that popjectification is conceptually prior to enriched engagement.

Reflective self-awareness does not preclude pre-reflective selfhood. The separated creature will continue to embody the simpler sorts of self-consciousness that it will have mastered en route to becoming fully separated. The creature will experience affect, effort, and bodily sensations. It will experience its body as physically coherent and bounded, though not always by reflecting on it *as* its body. It will be able to track its own spatiotemporal route through the world (Butterworth, 1995). Being inherently material, the creature will never totally disengage from the world, and will always be an unreflective ecological subject as well as a substantive

self. Even if it could be purely detached, by doing so it would cease to have the right sorts of interactions for understanding itself and others. Substantive self-consciousness is an added ability, not an usurpation of engaged self-perception (Smith, 1996; Legrand, 2007). We are at once conceptual selves and ecological selves, subjects and objects. A substantive self-conscious creature conceives of itself as a certain sort of object, a *person*, and a person is no more a disembodied consciousness than it is a lifeless hunk of matter; it is constitutively both at once. A sense of personhood is only achieved through the cooperation of these two aspects, the creature's physical reality and its mentality (Strawson, 1959). That is what makes conceptual awareness of one's personhood a landmark achievement for a thinking creature.

6: Objections and Questions Arising

Before concluding, I will tackle potential objections to the claims made throughout this paper. I have been arguing that the functions of visual perception enable new kinds of mental capacities, those of separation and engagement. These skills turn out to be deeply entwined, and each supports a different facet of self-consciousness.

A key question relates to the centrality of vision to this account. Granting that perception plays a crucial part in separation and engagement, there is still the further contention that of all the sensory modalities, it is vision which underlies these skills. Is the insistence on a visual foundation justified? There are certainly points that recommend vision over the other senses, at least in primate cognition. Visual streams carry out two vital parsings of the world. This functional dedication is more obvious in vision than other modalities. Moreover, visual processing areas comprise about sixty percent of the primate neocortex, and the human retina can register the smallest physical quantity of light that exists (Goodale, 1995). In rubber hand experiments, subjects are tricked into perceiving fallacious hand motions, despite contradictory proprioceptive information, by watching strategically placed dummy hands move in the wrong direction. Experiments like these demonstrate that vision can override much of our sensory perception, even 'private' sensations. As Jeannerod puts it, "vision has a prevalent role over other senses in [self-identification]: we feel our hand where we see it, not the converse" (2006, p. 73).

This first response is unsatisfying, though. It only shows that if visual function underlies separation and engagement, we have a contingently lucky pairing. At least two sizeable worries remain. Could other sensory modalities, alone or in cooperation, underlie separation and engagement? That is, is there anything more than a contingent relationship between vision and the abovementioned capacities? And, second, even given that vision enabled greater cognitive abilities in primate evolution, could it not be that the burgeoning evolutionary advantage afforded by separation and engagement drove the functional division within duplex vision? In other words, is the story backwards?

It is not unimaginable, of course, that there could be a non-visual duplex sensory modality interpreting the same world in two interlaced ways, perceiving parameters for action while stabilising the world for conceptual contemplation. But whether the scenario is metaphysically feasible is trickier. It is not the neurofunctional potential of other sensory modalities that is the problem, but rather how they relate to the properties of the material world. The material world is spatiotemporal, *profoundly* spatiotemporal. "[R]eal space," Grush remarks, "is where one's sensory states and behavioral actions are appropriately integrated" (2007, ms. p. 16). Engaged activity takes place in egocentrically constrained spaces, in real time, and separated thought is potent precisely because it defies local space, and because it time travels. Concepts refer instantaneously to objects in the farthest reaches of the galaxy, and these referents are seldom momentary time-slices (Smith, 1996). Objects are temporally and spatially extended, and conceptual thought captures some of this breadth. Separation is defined by its defiance of spatiotemporal dimensions, but only because space and time dictate its boundaries.

Where other senses may resolve temporal change, vision is the primary modality for resolving the world into spatiotemporal information. Touch, too, resolves sensation spatially, but only insofar as the sensations are smotheringly engaged. The body, being the spatiotemporal canvas for the sense of touch, brings perception of the world to a bottleneck at the skin-world barrier. Tactile perception could only satisfactorily depict the spatial world insofar as it brought a visual panorama to the skin's surface, as occurs in tactile-visual substitution systems (Grush, 2007). The difference is not one of the quantity of space accessible to a sensory modality; it is one of respecting all the different relational arrangements of objects that can occur in three dimensions. Touch is very limited in this regard; if a small cube presses against

my skin, the large globe behind it will be excluded from tactile perception, though not occluded visually. As Strawson's (1959) poignant Sound World thought experiment made clear, an auditory world, being non-spatial, cannot be as elaborate as the visual world in parsing raw information into stable particulars. The thought experiment stumbles on the need to endow the Sound World inhabitant with an auditory body rather than representing him as a non-extended perspectival origin, an indexical singularity (Strawson, 1959). Taste and smell have the same shortcomings as audition. Only in space and time does it makes sense to speak of causality, reidentification, engagement, separation, social interaction, and so on. That is why vision turns out to be the most suitable sense out of the five that humans have at their disposal.[8]

The second concern was that the relationship between vision, separation, and engagement is backwards. Duplex vision is not *really* behind separation and engagement; natural selection for the skills of separation and engagement drove the visual system to arrange itself to maximise functional dedication to these abilities. There are two ways of reading this objection. If it implies that evolution favoured visual systems which displayed signs of useful burgeoning capacities for separation and engagement, then it is not an objection at all. However, the objection becomes more interesting if it really means that other pre-existing separation and engagement capacities in the creature fostered the functional division in the visual brain.

On this reading, the objection, in one sense, is exactly right. There were certainly engaged activities the pre-visual or pre-duplex-visioned creature could partake in. It could *move about*, for one thing. But separation is, intuitively, a different matter. To deem that evolution could create a (proto-)separated creature which could not yet see is to make the strong claim that objects could be conceived of without the aid of an intrinsically spatial sensory modality. I have argued that there is much more to objects than their appearances in visual scenes. But I have also argued that these other characteristics must be tied to a coherent candidate which possesses them. I advanced popjects for this role. Could there be other candidates? An object concept would work, if it were not the case that evolution offers ample examples of nonconceptual sighted species and none of conceptual blind ones. Similarly, we might

---

[8] It is not the only metaphysically viable choice, though. As alluded to, tactile-visual substitution systems ought to enable both separation and engagement. A modality like echolocation might also suffice.

imagine that a language-user's separated thinking would make it advantageous to have a ventral stream that packaged the world into referents, were it not so strange to think that language could drive the evolution of vision. In our evolutionary history, at least, it is safe to say that separation skills were acquired downstream of becoming sighted. Unless we admit our sightless primeval ancestors among conceptual creatures, a less than compelling move (Carruthers, 2004), we must concede that sight emerged before separated thought. Whether this *had* to be the case is a matter of speculation about what other vehicles for objecthood could play the role of popjects.

There is in any case an alternative story about the bifurcation of the streams. Researchers claim that the architectural split in mammalian visual neurology was due to competing computational demands. Ventral and dorsal processing extract different information from scenes. The measures and invariants collected for seeing popjects and taking action are often incongruous; this would have led to conflicting demands on the brain that made it advantageous for a divided layout to emerge (Goodale & Milner, 2004). This proposal, though it draws on the conventional anatomical distinction, profits from its conciliatory tone. For in accepting it, we may regard that computational differences between processes contingently linked to separation and engagement made one midbrain visual system into two cortical ones, but that it was the subsequent real-time, real-world functioning of the two streams that allowed separation and engagement to emerge. Recourse to computational demands also casts it as a matter of mere complexity that separation and engagement have distinct neural backdrops. Since the streams' intertwined processing and the cooperation of separation with engagement are so critical for our embedded, embodied, and mental experiences, we should not expect the architectural division to be a matter of metaphysical significance, as if separation and engagement would have somehow tainted one another. It is exactly the reverse: neither skill can exist in anything but a stunted form without the other.

Onto questions about how all this ties in with self-consciousness. One might wonder whether, for instance, there could be a substantively self-conscious creature who could not see objects and who, therefore, lacked the capacity to detach itself from its environment. A caveat, first. It has been assumed throughout this discussion that modern-day human beings have cultural environments (and, perhaps, brains) that reinforce the development of separation and rich engagement to such an extent that visual pathologies ought not radically stunt the achievement of separated self-

consciousness. That is, we live in a post-separation society. Whereas during our evolutionary development duplex vision was the best faculty we had for distancing ourselves from our environments even as we acted within them, we now have language, and human infants' imitation capacities are now confronted with conspecifics who have already achieved general thought, producing a lot more fuel for separation than visible facial expressions. An evolving creature would not have had the support of such Gregorian tools for thinking. Contemporary patients like D.F. are substantively self-conscious, and optic ataxics are too, and that should come as no surprise.

Modern humans with visual brain pathologies are like aircrafts whose take-off thrusters have failed after they are already in the air. But what about candidates who are still on the runway, so to speak? I have already considered the issue of whether non-visual creatures could become separated. I concluded that other sensory modalities might do the job if they were inherently spatial, which is equivalent to a claim that non-visual creatures might achieve substantive self-consciousness. But these creatures do not bypass separation, they merely take a different route through it. There is still the question of how tight the link is between separation and substantive self-consciousness.

Could substantive self-consciousness arise in creatures unable to think about objects, and about what is spatially or temporally removed? Is separation necessary for awareness of oneself as an object in the world? I believe the answer is yes, on *a priori* grounds. To think about oneself as an object, one must think of objects. To think of objects just is the capacity to think about what is far away and temporally disconnected insofar as even thinking about objects in a visually present scene (as opposed to popjects in a scene) is thinking of them as possibly being elsewhere, or moving to different locations, where they will be potentially reidentified (as opposed to recognised) by the observer. A creature which did not contemplate objects would never pull itself free of closed, causally effective loops with its environment (Smith, 1996). Creatures like Brooks' (1991) *mobots*, if they do not see thick objects, only affordances for action, will never become substantively self-conscious. Similarly, a fully separate mind with no engagement abilities would never achieve substantive self-consciousness either. A von Neumann computer, given its lack of world-directed drive, action understanding, imitation skills, and grasp of causal regularities, would fall short of objectifying itself in thought.

Finally, one might ask what, if anything, distinguishes human self-consciousness from that of animals which have duplex visual systems. Are we willing to ascribe separation and substantive self-consciousness to cats and dogs? We might wish to do so, if we are willing to credit them with some kind of conceptuality. Pets may be substantively self-conscious but pre-linguistic. Language may just tip the scales of intelligence, and the outward appearance of mindedness, that crucial bit further, as Dennett claims (1996). After all, humans are not just substantively self-conscious. We place ourselves within self-narratives and contemplate our positions in wider histories. This talent might look surprisingly more impressive from the outside than basic substantive self-consciousness does.

Let us suppose, however, that cats are not substantively self-conscious, being creatures unable to think of themselves as objects and manipulate those thoughts without hindrance from the world as it presents itself to them. Why would this be the case? As I have just said, separation is necessary for substantive self-consciousness. Perhaps such creatures are not properly separated from their perceptual worlds. This could be for a number of reasons. Empirical findings could narrow down the possibilities. Cats might have visuomotor systems which do not support imitation abilities or which otherwise place limits on motor cognition. Similarly, their ventral streams, or the interconnections between their ventral and dorsal streams, may be less elaborate than ours. It would be revealing if animals were found not to perceive dorsal affordances in designed objects because they could not ventrally recognise them as tools, resulting in oblivious visually guided action towards certain items. Neither is duplex vision a sufficient neural basis for full-fledged separation. Some animals might not have the attention or memory capacities to reidentify objects or even sense familiarity. Rather, they may display raw stimulus-response behaviour towards 'recognised' items or mechanically habituate to repeated stimuli.

More likely, however, is that animals' *brains* are fine. It is just that animals do not form social groupings that are conducive to separated thought. "[O]ur kind of thinking," claims Dennett, "had to wait for talking to emerge, which in turn had to wait for secret keeping to emerge, which in turn had to wait for the right complexification of the behavioral environment" (1996, p. 172). Secret keeping requires animal groupings that are collaborative but which allow each individual to wander off, alone, and discover environmental advantages that it might not wish to share. Domesticated animals, as intelligent as they seem, tend to lead solitary lives at

a distance from conspecifics, whereas wild animals tend to confine themselves to crowded, cooperative groupings.

7: Concluding Remarks

In this essay, I have attempted to show that the functions of the two visual streams enable certain embedded cognitive skills which together permit a sophisticated kind of self-consciousness to develop. The visual streams in question are ventral vision-for-judgement and dorsal vision-for-action. These cognitive skills, I have termed *separation* and *engagement*. Ventral and dorsal vision are necessary for separation and rich engagement and are the main enabling factors – contributions from memory and attention systems aside – for proto-separation and modest engagement. In turn, proto-separation and engagement are necessary and, when intertwined and undertaken in cultural surrounds, sufficient for full-fledged separation, which is the world-based skill that maps onto the mind-based capacity for general, conceptual thought. (Conceptual mastery of abstract concepts may require some further cognitive sophistication than separation.) Separation, with contributions from proto-separation and visuomotor understanding, and in the context of continued enactive experience, is necessary and sufficient for substantive self-consciousness, the capacity to think of oneself as an object. In this way, the account sketched here has risen from dichotomies at the neurological level to an integrative, situated recipe for selfhood at the psychological level.

Questions remain unanswered, and avenues have been left unexplored. Most compelling are considerations about the way that subjectivity and objectivity make contact with the present concerns. Does being a subject more closely parallel the ecological self or the substantive self? Does objectivity come hand in hand with separation, or does it require further sophistication? Perhaps it is a prior accomplishment of the proto-separated creature. I suspect that the answers to these questions will require continued integration between thought and action in the world. It would also be helpful to more exhaustively enumerate the factors that enable full-fledged separation and the transition from seeing popjects to thinking about objects. The selection discussed in this paper was intended to highlight the indispensability of dorsal contributions to separation, and may have neglected several major conditions

on becoming a separated thinker. Specifically, the roles of language and culture deserve in-depth analysis.

Finally, an interesting question is how seeing popjects and thinking about objects relates to what is true. There is something appealing about Smith's (1996) idea that what is taken to exist is both a human construction and normatively governed by the world, so that it makes sense to say that some carvings up of the world into objects do violence to it, while others respect its constraints. But if that is right – and if so, it probably percolates down into the flexibility inherent in perceptually selecting popjects – then are the conceptual thoughts which better respect the world the true ones, or is this conclusion a misleading temptation? *Something* pins the warrant for our separated thoughts, floating free of time and space, back down to the bedrock material world, and whether this is reference, truth, coherence, or simply something like living ethically and truly (Smith, 2001), remains to be seen.

Bibliography

Avila-García, Orlando & Cañamero, Lola (2005). *Hormonal Modulation of Perception in Motivation-Based Action Selection Architectures*. Paper presented at the AISB'05 Symposium: "Agents That Want and Like: Motivational and emotional roots of cognition and action", University of Hertfordshire, United Kingdom.

Bedford, Felice L. (1997). False Categories in Cognition: The Not-The-Liver fallacy. *Cognition*, *64*, 231-248.

Bermúdez, José L. (1995). Ecological Perception and the Notion of a Nonconceptual Point of View. In J. L. Bermúdez, A. Marcel, & N. Eilan (Eds.), *The Body and the Self*. (pp. 153-173). Massachusetts: Bradford Books, MIT Press.

Bermúdez, José L. (2007). From Two Visual Systems to Two Forms of Content? *Psyche*, *13*(2), 1-7.

Brewer, Bill (1995). Bodily Awareness and the Self. In J. L. Bermúdez, A. Marcel, & N. Eilan (Eds.), *The Body and the Self*. (pp. 291-309). Massachusetts: Bradford Books, MIT Press.

Brooks, Rodney A. (1991). Intelligence Without Representation. *Artificial Intelligence*, *47*, 139-159.

Butterworth, George (1995). An Ecological Perspective on the Origins of Self. In J. L. Bermúdez, A. Marcel, & N. Eilan (Eds.), *The Body and the Self*. (pp. 87-105). Massachusetts: Bradford Books, MIT Press.

Campbell, John (1993). The Role of Physical Objects in Spatial Thinking. In N. Eilan, R. A. McCarthy, & M. W. Brewer (Eds.), *Spatial Representation: Problems in philosophy and psychology*. (pp. 65-95). Oxford: Basil Blackwell.

Campbell, John (1995). The Body Image and Self-Consciousness. In J. L. Bermúdez, A. Marcel, & N. Eilan (Eds.), *The Body and the Self*. (pp. 29-42). Massachusetts: Bradford Books, MIT Press.

Campbell, John (2006). What is the Role of Location in the Sense of a Visual Demonstrative? Reply to Matthen. *Philosophical Studies*, *127*, 239-254.

Carruthers, Peter (2004). On Being Simple Minded. *American Philosophical Quarterly*, *41*, 205-220.

Carruthers, Peter (2005). Why the Question of Animal Consciousness Might Not Matter Very Much. *Philosophical Psychology*, *18:1*, 83-102.

Chemero, Anthony (2007, *forthcoming*). Situated, Embodied Realism. In J. Burgos & E. Ribes (Eds.), *Knowledge, Cognition and Behavior*.

Clark, Andy (2002). That Special Something: Dennett on the making of minds and selves. In A. Brook & D. Ross (Eds.), *Daniel Dennett*. (pp.187-205). Cambridge: Cambridge University Press.

Clark, Andy (2007). What Reaching Teaches: Consciousness, control, and the inner zombie. Retrieved June 25th, 2007, from http://www.philosophy.ed.ac.uk/contact/draftpapers.html

Clark, Andy & Chalmers, David J. (1998). The Extended Mind. *Analysis*, *58*(1), 7-19.

Clark, Andy & Grush, Rick (1999). Towards a Cognitive Robotics. *Adaptive Behavior*, *7*(1), 5-16.

Clark, Andy & Toribio, Josefa (1994). Doing Without Representing? *Synthese*, *101*, 401-431.

Clark, Austen (2004). Feature-Placing and Proto-Objects. *Philosophical Psychology*, *17*(4), 443-469.

Cohen, Yale A. & Andersen, Richard A. (2002). A Common Reference Frame for Movement Plans in the Posterior Parietal Cortex. *Nature Reviews Neuroscience*, *3*, 553-562.

Cussins, Adrian (1992). Content, Embodiment and Objectivity: The theory of cognitive trails. *Mind*, *101*(404), 651-688.

Dennett, Daniel C. (1996). *Kinds of Minds*. New York: Basic Books.

Eilan, Naomi (1995). Consciousness and the Self. In J. L. Bermúdez, A. Marcel, & N. Eilan (Eds.), *The Body and the Self*. (pp. 337-357). Massachusetts: Bradford Books, MIT Press.

Eilan, Naomi; Marcel, Anthony; & Bermúdez, José L. (1995). Self-Consciousness and the Body: An interdisciplinary introduction. In J. L. Bermúdez, A. Marcel, & N. Eilan (Eds.), *The Body and the Self*. (pp. 1-28). Massachusetts: Bradford Books, MIT Press.

Evans, Gareth (1982). *The Varieties of Reference*. Oxford: Oxford University Press.

Frege, Gottlob (1980). *The Foundations of Arithmetic*. Second Edition, J.L. Austen (Trans.). Illinois: Northwestern University Press.

Gallese, Vittorio (2007). The 'Conscious' Dorsal Stream: Embodied simulation and its role in space and action conscious awareness. *Psyche*, *13*(1), 1-20.

Gallese, Vittorio & Lakoff, George (2005). The Brain's Concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3/4), 455-479.

Gibson, James J. (1979). *The Ecological Approach to Visual Perception.* New Jersey: Erlbaum.

Gillett, G. R. (1987). The Generality Constraint and Conscious Thought. *Analysis, 47*(1), 20-24.

Glover, Scott (2004). Separate Visual Representations in the Planning and Control of Action. *Behavioral and Brain Sciences, 27*, 3-78.

Goodale, Melvyn A. (1995). The Cortical Organisation of Visual Perception and Visuomotor Control. In S. Kosslyn & D. Osherson (Eds.), *Visual Cognition: An invitation to Cognitive Science*, Second Edition. (pp. 167-213). Massachusetts: MIT Press.

Goodale, Melvyn A. (2005). Action Insight: The role of the dorsal stream in the perception of grasping. *Neuron, 47*(3), 328-329.

Goodale, Melvyn A. & Milner, David (2004). *Sight Unseen: An exploration of conscious and unconscious vision.* Oxford: Oxford University Press.

Goodale, Melvyn A. & Westwood, David A. (2004). An Evolving View of Duplex Vision: Separate but interacting cortical pathways for perception and action. *Current Opinion in Neurobiology, 14*, 203-211.

Grill-Spector, Kalanit (2003). The Neural Basis of Object Perception. *Current Opinion in Neurobiology, 13*, 159-166.

Grush, Rick (2007, *draft*). Manifolds, Coordinations, Imagination, Objectivity.

Hatfield, Gary (1990). Gibsonian Representations and Connectionist Symbol Processing: Prospects for unification. *Psychological Research, 52*(2-3), 243-252.

Jeannerod, Marc (2006). *Motor Cognition: What actions tell the self.* Oxford: Oxford University Press.

Jeannerod, Marc & Jacob, P. (2005). Visual Cognition: A new look at the two-visual systems model. *Neuropsychologia, 43*, 301-312.

Johnson, Helen & Haggard, Patrick (2005). Motor Awareness Without Perceptual Awareness. *Neuropsychologia, 43*, 227-237.

Jordan, J. Scott (2003). Emergence of Self and Other in Perception and Action: An event-control approach. *Consciousness and Cognition, 12*, 633-646.

Kirsh, David (1991). Today the Earwig, Tomorrow Man. *Artificial Intelligence, 47*, 161-184.

Legrand, Dorothée (2007). Pre-Reflective Self-Consciousness: On being bodily in the world. *Janus Head, 9*(2), 493-519.

Loomis, Jack M. (1992). Distal Attribution and Presence. *Presence: Teleoperators and Virtual Environments*, *1*, 113-119.

Matthen, Mohan (2005). *Seeing, Doing, and Knowing*. Oxford: Oxford University Press.

McIntosh, R. D.; McClements, K. I.; Schindler, I.; Cassidy, T. P.; Birchall, D.; & Milner, A. D. (2004). Avoidance of Obstacles in the Absence of Visual Awareness. *Proceedings of the Royal Society*, *271*, 15–20.

Meltzoff, Andrew N. & Moore, M. Keith (1995). Infants' Understanding of People and Things: From body imitation to folk psychology. In J. L. Bermúdez, A. Marcel, & N. Eilan (Eds.), *The Body and the Self*. (pp. 43-69). Massachusetts: Bradford Books, MIT Press.

Norman, Joel (2002). Two Visual Systems and Two Theories of Perception: An attempt to reconcile the constructivist and ecological approaches. *Behavioral and Brain Sciences*, *24*(6), 73-96.

Nudds, Matthew (2007, *draft*). Seeing How to Move: Visually guided action and the 'directive' content of visual experience.

Perry, John (1979). The Problem of the Essential Indexical. *Noûs*, *13*, 3-21.

Prinz, Jesse J. (2002). *Furnishing the Mind: Concepts and their perceptual basis*. Cambridge: MIT Press.

Prinz, Jesse J. (2007, *forthcoming*). The Return of Concept Empiricism. In H. Cohen & C. Leferbvre (Eds.), *Categorization and Cognitive Science*. Elsevier Press.

Prinz, Wolfgang (2003). Emerging Selves: Representational foundations of subjectivity. *Consciousness and Cognition*, *12*, 515-528.

Proust, Joëlle (2003). Thinking of Oneself as the Same. *Consciousness and Cognition*, *12*, 495-509.

Pylyshyn, Zenon W. (2000). Situating Vision in the World. *Trends in Cognitive Sciences*, *4*(5), 197-207.

Pylyshyn, Zenon W. (2001). Visual Indexes, Preconceptual Objects, and Situated Vision. *Cognition*, *80*, 127-158.

Rossetti, Yves; Pisella, Laure; & Vighetto, Alain (2003). Optic Ataxia Revisited: Visually guided action versus immediate visuomotor control. *Experimental Brain Research*, *153*, 171-179.

Schenk, Thomas (2006). An Allocentric Rather than Perceptual Deficit in Patient D.F. *Nature Neuroscience*, *9*(11), 1369-1370.

Smith, Brian Cantwell (1996). *On the Origin of Objects.* Massachusetts: Bradford Books, MIT Press.

Smith, Brian Cantwell (1999). *The Nonconceptual World*. Paper presented at Duke University.

Smith, Brian Cantwell (2001). *Subjectivity and Objectivity*. Paper presented at the Thomas Langford Inaugural Lecture, Duke University.

Stern, Daniel N. (1985). *The Interpersonal World of the Infant*. New York: Basic Books.

Strawson, Peter F. (1959). *Individuals: An essay in descriptive metaphysics.* London: Methuen & Co. Ltd.

Westwood, David A.; Chapman, Chas D.; & Roy, Eric A. (2000). Pantomimed Actions May Be Controlled by the Ventral Visual Stream. *Experimental Brain Research*, *130*, 545-548.

Zahavi, Dan (2002). First-Person Thoughts and Embodied Self-Awareness: Some reflections on the relation between recent analytical philosophy and phenomenology. *Phenomenology and the Cognitive Sciences*, *1*, 7-26.