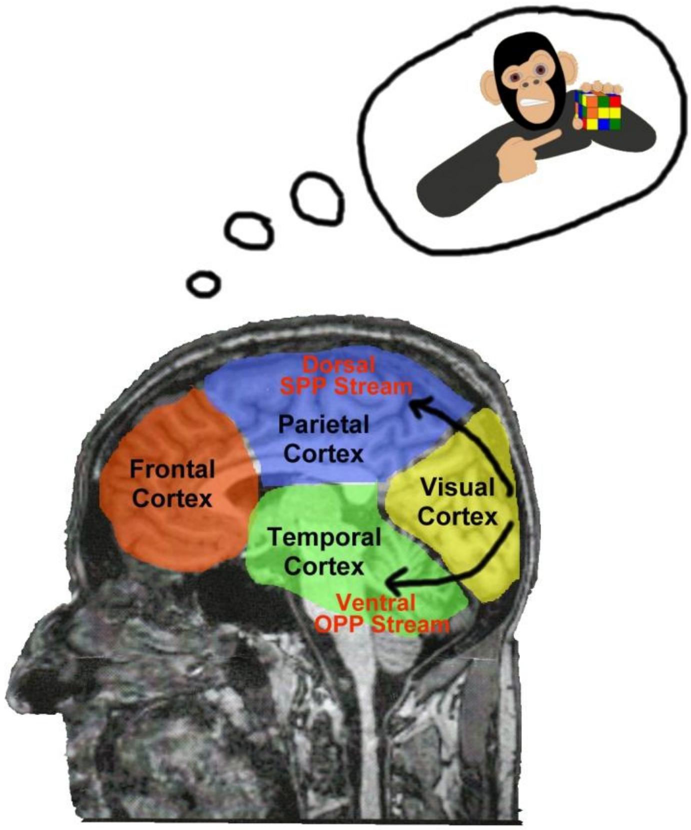
What is the Nature and Function of Spatio-Temporal Imagination?

Can it plausibly be explained as an offline simulation of the visual process?

Peter R. King BSc, MSc, BA, MA dic.

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

September 2009



Abstract

The main aim of this thesis is to analyse the nature and function of Spatio-Temporal Imagination (STIm) under an imagination-as-simulation framework. STIm is defined as *any imaginative act that allows you to consider a location and/or a time other than the one you are currently experiencing*. I will focus on the phenomenal visual imagery aspect of this mental phenomenon and hence will not discuss other potential versions of STIm in any detail.

Part I will explore the current simulation literature to get an idea of what imagination-as-simulation may mean and to see how it dovetails with other researchers who have a similar style of approach (Chapter 2). I will then defend the Kosslyn Model of visual imagery, which I will argue is ameneable to being interpreted as an imagination-as-simulation process (Chapters 3&4).

Part II will look at what different perceptual theories may say about the imagination-as-simulation debate. Using the Selective vs. Generative Approach distinction as a guide, we will look at two theories that are obviously examples of each respectively: Naïve Realism (Chapter 5) and my own theory of STIm (Chapter 7). We will also look at Tye's (1995) PANIC theory, as a related approach that will also be used to illustrate other points relevant to the debate about the relation between perception and imagery (Chapter 6).

I will argue that visual STIm has much in common with visual perception in that they are both: *locally generated visual style phenomenal representational mental states*. This is even though there are some other crucial differences in how they are caused and experienced. The thesis will also describe a way that these images can be labeled with different spatial and temporal contexts. This is what allows STIm to be used to consider alternative possibilities both temporally and spatially and to function as a way to plan our actions in the present and to have an extended spatial and temporal awareness of our environment.

'Imagery is like a juggling act in which only a small number of balls can be kept aloft at once.' Kosslyn et al. (2006: 42)

Acknowledgments

I would first of all like to thank my supervisor Professor Gregory Currie for his advice, guidance and inspiration during my PhD and previous studies. I would also like to thank Mathew Kennedy and Bill Fish for their feedback and helpful comments, especially Bill's help towards the end. I would also like to thank various other people who attended my talks at Nottingham on different parts of the thesis and especially Ben Smart for reading and commenting on some sections for me. I would also like to thank my family, and especially my parents, for their continued support. I would like to thank Oscar and Molly for being smashing and my juggling friends for keeping me sane. However by far my biggest debt of gratitude is to my partner Cindy-Lee Tickets who put up with my ups and downs and general neglect of my personal life. I thank her for her patience and loving support and I dedicate this thesis to her and my late Grandpa and Grandma.

Cover picture: Scan of author's brain and representation of what he's been thinking about and where (see also Appendix B).

CONTENTS

1 - Introduction

> 'My main aim is to demonstrate that imagination is a faculty that runs through the most diverse of mental phenomena; it is a theme on which there are many variations. We need imagination to have mental images, to dream, to believe, to represent possibilities, to mean. …the imagination, I contend, is a far more pervasive aspect of the human mind than recent philosophy might suggest.' Colin McGinn (2004: 5)

1.1 - What is Spatio-Temporal Imagination (STIm)?

Imagination is a broad topic and it can be categorised in a number of different ways. For instance you can divide it up according to whether it has a phenomenal 'what it is like' aspect to it (Nagel 1974) and this could be labelled phenomenal or sensory imagination. If it lacks this sensory aspect then it may merely involve considering a more semantically based counterfactual, such as: your birth date being 20 years later; or what would happen if the president died. You may want to refer to this as supposition and it could be categorised as propositional or semantic imagination. This initial supposition may then lead to some further, more obviously phenomenal, imagery and the generation of other associated memories. It may even lead to an extended daydreaming episode that surprises you where you end up, for example: imagining yourself as a young vice-president who gets a sudden promotion. It may also involve aural imagination as you go over the options using the 'voice in your head'. You can also explore imagination based on what it does, or how it works, and it may be hard to rigorously isolate one category from another. This may be because these categories overlap somewhat and that they work quite closely together. The use of imagination to order our thoughts is probably more pervasive and inter-linked than at first you might have suspected (and see opening quote).

This thesis involves an analysis of a sub-category of imagination that I call Spatio-Temporal Imagination (STIm). My definition of STIm is *any imaginative act that allows you to consider a location and/or a time other than the one you are currently experiencing*. This will centrally include imagining a future possible situation or considering a past counterfactual situation. But it may just involve imagining a different location in the present. There are a few odd cases of STIm where you may be imagining the local area in the present slightly differently; or even more oddly, the present as it is. You may also just be imagining an object (e.g. a camel) in abstraction from any spatial or temporal location in order to work something out (e.g. what shape ears it has). For the purposes of this essay I will treat these as special cases of imagining spatially and where the exact location or temporal context is left unspecified. It should be obvious from the thesis that they can certainly be explained in the same way as other more central cases of STIm, without having to invoke any new mechanisms or capacities. At any rate it will become clear as we progress how these different varieties can be accommodated by my theory. This is because in order to develop a robust theory of STIm, I will also develop a general theory of imagery.

So even though there are many ways to categorise imagination, I do not propose to undertake an extensive analysis of the imagination literature. In this thesis I will try to keep the focus as narrow as possible and hence restrict my analysis of STIm in two ways. First of all, it is already restricted by content, or what it is about, in that it allows us to consider another time and/or place. Secondly I will restrict my analysis to vision-like phenomenal imagination of the kind often used in visualization and hence I will not address non-sensory versions of STIm. I class something as *vision-like* if it is seemingly experienced in a *visual style* that involves features like apparent shapes and colours in a similar way that visual experience does. This is even though these features may only be apparent and they don't actually posses actual 3D shape and colour. The vision-like category definitely includes visual hallucinations, after-images and visual imagery, but may include other vision-like states that I do not mention here. I use the term *visual style* to indicate an experience that occurs in a modality that features apparent or real shapes and colours and this includes both vision and vision-like mental states.

Because of the restricted focus just described, I will spend some time looking at visual mental imagery that allows us to phenomenally imagine a different time and/or place. I will also look at the mechanisms that underlie this phenomenon and examine whether visual imagination can best be seen as an offline version of perception. That is, we will look at how closely visual imagination is linked to visual perception and whether some of the mechanisms that are used in perception, are also used in an offline (or mock, or pretend) manner in visual imagination. I call this an *imagination-as-simulation* approach and it can be seen in the wider context of other simulationist approaches to the mind: e.g. the simulation theory of mindreading. There is some initial evidence that suggests many of the brain areas active during perception are also active during visualization and this is initially indicative that imagination may be an offline simulation of perception[1]. We will look very closely at what this might mean in what follows.

1.1.1 – STIm in general and links to other subject areas

Before we do that, here are some central examples of STIm in action, which gives an initial indication of the wider function I think it may serve in our mental economy. I have appended each one with its possible temporal and spatial aspects in brackets:

[1] Cf. Kosslyn et al (1997 & 2001), Currie (1995a,b), Currie and Ravenscroft (1997, 2002) and Goldman (2006).

- Choosing between long term plans in life by forming an image of a future situation and assessing your emotional reaction to it to decide if you want to work towards that goal or not. Then working out the steps you believe are necessary to achieve that goal, which may in turn also involve imagining these steps in detail (distant to nearer future / variable location).
- Thinking about what you want (desire) to do next and how to achieve it (belief). Perhaps by imagining how the room you are in needs to be changed and how it may look when you change it (near future / local).
- Listening to someone explain something that is (or has been, or will be, or could be) going on elsewhere, and generating an idea of what that may look like in your imagination (current, past or future / elsewhere).
- Thinking about how you could have done something else in the past and how this may have made the present work out differently. Perhaps then using this to learn from your mistakes and prevent the same thing happening again. Conversely you may want to repeat good practices that turned out well (past counterfactual or factual / variable location).
- Working out how to pack the trunk of your car most efficiently when you are off on a camping holiday. You may need to visualize which bags go best where and even mentally rotate them to try out different combinations before ever lifting a bag. (future / local).
- Solving the Tower of Hanoi puzzle by keeping track of the moves you have already done and anticipating moves you may need to make in the future. Keeping the final goal in mind and judging whether each move is helping you to achieve it (future, past/ local). [NB. See Appendix A for an explanation of the Tower of Hanoi puzzle]

The main idea of giving these examples is to suggest that STIm plays a certain role amongst many other mental states. These include: emotions, beliefs, desires (or goals), memories, language comprehension etc. Solving the Tower of Hanoi puzzle is used as a test of executive function deficit and failure to solve it may be indicative of damage to the prefrontal cortex or of mental disorders such as autism and schizophrenia[2]. It also indicates how it may be crucially involved in supporting your personal identity over time and people who lack it may be stuck in the present without any personal history or future aspirations[3]. Talking like this also parallels the new and expanding Episodic Thinking and Mental Time Travel (MTT) literature[4]. The former involves thinking using episodes from your past life or predicting possible episodes in your future life, and the latter involves mental projections to other times. I think that STIm can certainly be seen as a form of episodic thought and certainly involves mental projection to another time and indeed much of this analysis has been inspired by that literature and is aimed to eventually contribute and comment on it. Consider the table below as a way of illustrating the way I see STIm possibly matching up with this growing field:

---

[2] C.f. Currie (1996), Carruthers (1996), Boucher (1996), Perner and Lang (2000), Pennington & Ozonoff (1996) and Russell (1997).
[3] C.f. Subject K.C. described in Tulving (2002).
[4] C.f. Attance & O'Neill (2005), Clayton et al (2003), Suddendorf and Corballis (1997, 2007).

EPISODIC THINKING SYSTEM		
MTT to Past		**MTT to Future**
Spatio-temporal Memory	**STIm**	
Actual / Factual	*Counterfactual / Possible*	*Possible*
Episodic Memory	Episodic Imagination	Episodic Imagination

Figure 1.1 – One possible way to represent the relationship between
Episodic Thinking, Mental Time Travel and STIm

Unfortunately it will not be possible to develop these other themes in detail here, however I will mention a little bit more about them as we go along and this will help illustrate what I see as the main function of STIm. For now the main idea is to show how explicit visualization of a non-actual situation, and making judgments about it, can help you plan and guide your actions in the present. It also seems possible to suggest that STIm had a central role to play in strategy testing in an ancestral environment and hence would greatly increase our survival and reproductive chances. This is perhaps how it has earned its evolutionary keep and why it may have come into existence in the first place. I have written on this aspect elsewhere[5] and will mention it at times as we progress. Gregory Currie has suggested that it is this evolved capacity that we access today and may be the basis for certain other mental phenomena:

> 'Strategy testing is, on my hypothesis, the proper function of the imagination; the function appeal to which explains why we have the faculty of imagination. …Daydreaming and fantasy, along with imaginative involvement in fictions, are made possible by a system that already exists for other purposes: strategy testing.'
> Currie (1995b: 158)

[5] C.f. King (2003, 2005a,b,c).

This ability to strategy test is perhaps currently most extremely exhibited in the imaginative ways that we seem to find of anticipating each other's actions and killing each other in the context of warfare and espionage.

Related to that point, perhaps another special situation involving STIm is where you may be imagining a different spatio-temporal situation that involves other creatures and their own decision making process. I refer to this as *Mentalizing Imagination* and it obviously links with the literature on Mind Reading and Theory of Mind[6]. I see this as another possible extended case of STIm, but where the objects involved in the imagination are now attributed intentional mental states of their own. This means you not only now have to predict what these objects will do, but you also have to predict this based on imagining what they are thinking, since this inevitably affects their actions. I see Mentalizing Imagination as a more developed and complicated version of STIm, requiring a greater intelligence and awareness of ones surroundings and the creatures that inhabit it. For the purposes of this thesis however we will stick to analysing STIm on its own, but I am aware of the wider implications of this thesis in this respect and I have investigated this elsewhere[7]. The above is aimed at illustrating the wider contexts that this thesis can be situated in and what the primary function of STIm might be.

1.1.2 – STIm: the broad structure of the thesis and background assumptions

Given the broad context above, I can now be more specific about the structure of this thesis. The main approach of this thesis can be split into two main streams. The **first** is to look at the mainly psychological literature on mental imagery and to develop a theory of the neural underpinnings of visualization. To that end I will focus on the visual imagery debate and I will defend a version of what I refer to as the 'Kosslyn Model' (Kosslyn et al

[6] C.f. Carruthers & Smith (1996), Davies & Stone (2000)
[7] Especially in my MA thesis on Stich and Nichols' (2003) book called 'Mindreading'.

2006 and see below). This defends a *hybrid depictive representational* approach, based on a 'visual buffer' in the visual cortex. As mentioned earlier, empirical evidence suggests that vision and visual imagery both utilise the visual buffer and that they compete for resources in the same sense modality. This immediately motivates an imagination-as-simulation approach, since we can argue that perhaps a good initial explanation of these phenomena is that the visual buffer is taken offline to support visual imagery. By offline I mean that the visual buffer is fed internally generated mock inputs, and these are treated in an 'as-if' or pretend way. Generally this means they will need to be *labelled* or *flagged* as mock inputs in some way during processing, so that they are not treated as normal inputs. The first half of the thesis (Chapters 2- 4) therefore deals with this more *cognitive processing* aspect of STIm, with Chapter 2 explaining what I mean by imagination-as-simulation in detail. Chapter 2 will also discuss ways that other researchers utilise simulation to develop their theories. For example: Grush's (2004) emulator theory; and Hurley's (2008) shared circuits model. Both of which appeal to simulation processes in some way. This will give us an idea of the further potential and utility of this approach, as well as giving us some further useful tools to use in the rest of the thesis: e.g. visual emulators. Chapter 3&4 will be concerned with explaining and defending the Kosslyn Model and I will say a bit more about this below.

The **second** half of the thesis looks at visual imagery from the other side of the signal to phenomenal gap and sees what certain perceptual theories have to say about imagery. This is motivated by the idea that if you are suggesting that imagery is an offline simulation of an online perceptual process, then it would be helpful to *know what that online process is thought to be doing in the first place*. If you have an idea of what is thought to be happening in the online case, then perhaps you can get an idea of what else needs to be done to achieve the offline imagistic version. It may be that feeding the visual buffer mock inputs is enough, or it may be that extra steps are required. Therefore the second half of the thesis (Chapter 5-7) will look at different perceptual theories and investigate their implications for my imagination-as-simulation approach. The rest of this introductory chapter will establish all the groundwork necessary before moving on to analyse each of these aspects in detail.

On a terminological note I will only use the phrase 'qualia' without qualification if I am referring to it in the narrow sense of being non-intentional, intrinsic properties of experience (c.f. Block 1996). I will use the more general term *phenomenal character* for the wider sense of 'qualia', that simply means any qualitative experience that has a what-it-is like aspect to it. It should also be noted that throughout this thesis I use the term 'phenomenal character' relatively non-technically, to merely describe the character of ones phenomenal experience, and not to particularly mean a property of an experience or anything else like that. Hence if you are experiencing a blue object, the phenomenal character of your experience is blueness; or to put it another way, your experience has the character of blueness which you experience phenomenally. This could be because the object you are directly experiencing is itself actually blue, or because your brain is putting you in an indirect perceptual state that represents to you that there is a 'blue' object in front of you. Hence the use of this term does not pre-suppose adopting any particular approach to perception. So phenomenal character, as it will be utilised here, just characterises the qualitative nature of your experience and makes no comment as to what this is ultimately a property of[8].

[8] But see Fish (2009: 6) for a discussion of its more narrow and technical use.

1.2 –Visualization, memory and after-images

Before we go on it may be a useful to get a stronger first person idea of what I mean by a visual image and perhaps a good way of doing this is to experience some paradigmatic examples of mental imagery yourself. To that end try to answer the following two questions:

1) How many small squares are there in total on all the faces of a Rubiks Cube puzzle?
2) How many windows are there in your house?

It may be helpful to take some time at this point to try to work these out and to take note of the kinds of mental steps involved. If you don't know what a Rubiks Cube[9] is then you can imaginatively construct an image of one as you read through the solution below and see how closely it matches the one depicted in Figure 1.2 overleaf.

1.2.1 - Visualization – the Rubiks Cube Problem

For the **first** question it is normal to form a mental image of a Rubiks Cube so as to be able to count off, or work out, how many small squares there are on each face (or at least this is what I will assume here and this will be discussed further in what follows). A Rubiks Cube has three rows of three squares per face, so that's nine per face. It may then be necessary to mentally rotate the cube and work out, or check, how many faces there are on a cube. It's possible you just know this from memory already. Then after some simple arithmetic (9x6) the solution is reached (54).

It is worth noting that the claim here is not that this is the only way to the answer, some people can do this by simply calling up the facts in a fairly non-sensory and semantic way.

[9] The actual term is 'Rubik's Cube': but I am just using 'Rubiks Cube' without the apostrophe to avoid textual clutter.

The claim here is that in most cases a phenomenal image of a Rubiks Cube is experienced and this helps with finding the answer. At the very least one might claim that it allows you to know how you know the answer, and perhaps gives you a certain access to how you have worked it out and how confident you are of the solution[10]. While some people deny they have any mental imagery at all, studies (Kosslyn et al 2006: 179) have shown that the majority of people claim to enjoy some form of phenomenal imagination. While I will discuss eliminativism about phenomenal imagination in Chapters 3&5, for the most part I will assume that there is such a thing and that it requires some kind of explanation. This thesis attempts to give one such possible explanation as far as I think is currently possible.

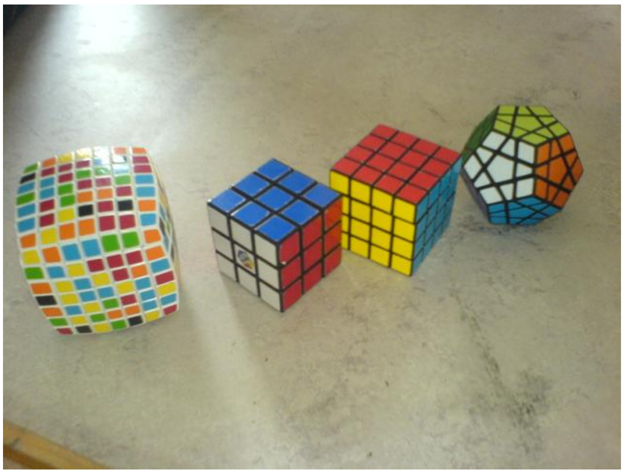


Figure. 1.2 – Examples of different kinds of Rubiks-style Cubes.
The Rubiks Cube referred to in the text is the second from the left (red, white, blue: with 27 of its small squares showing)

The difficulty of this explanatory task is illustrated by asking *what* this visual image of a Rubiks Cube *actually is* and where, if anywhere, it occurs. The usual thing to say at this point is that it is experienced in some 'inner imaginary space', but it is then very hard to get a tangible grip on what that is or to provide any coherent explanation of what inner imaginary space might be. What we certainly don't want to say is that there is an actual

[10] In other words it may help you understand how you know the answer and how confident you are of it by double-checking it. This is perhaps a form of meta-cognition (thoughts about thoughts) and may indicate another use for visual imagery (c.f. Rosenthal 2000 and other papers in that volume for discussions on meta-cognition).

Rubiks Cube in your brain, because if this is so, then you are probably in a lot of trouble! Neither do we want to say there is a literal picture of a Rubiks Cube in your head, which you see with an inner eye and rotate with inner hands, because this also leads to further problems explaining these features (cf. Block 1983). So some other form of explanation of this visual image may be appropriate here and this thesis is aimed at doing just that. This will also include explaining some of the following phenomenological observations about visual imagery:

1. It is harder to maintain and doesn't last very long in most cases.
2. It is generally less vivid and complete that an equivalent visual episode, although this can vary significantly.
3. It can be willed or occur spontaneously and it can contain surprising contents.
4. It can be manipulated and this can 'simulate' actual physical manipulation in certain ways e.g. rotation and scanning.
5. It is generally not experienced as being out there in the world, in contrast with afterimages, illusions, perceptions and a certain kinds of hallucinations (see below).
6. It generally seems to be limited in the same way that visual experience is: e.g. you imagine in shapes and colours similar to those in your potential perceptual range.

As the thesis proceeds all these features will be touched on and explained in more or less detail.

At this point the reader may think that the experiments by Perky (1910) may offer a counterexample to the suggestion above (i.e. entry number 5) that imagined objects are not experienced as if they are in the external world. In these experiments people were asked to imagine an object (e.g. a banana) whilst looking at a white screen. At the same time, and without their knowledge, that same object was faintly presented on the screen at just above the threshold of awareness. The subjects were reported to have been unable to tell these faint images apart from their imagery and sometimes were even surprised that they had imagined features that were in fact features of the faint image. This may indicate that imagery is experienced externally in a similar way as the faint projections. However these results were hard to duplicate by later experimenters (c.f. Segal 1971) and can be interpreted as due to confusing a perception of a projection for a self generated mental image, rather than making any predictions about our normal imagistic phenomenology.

In the after-images section (1.2.3) below I will give one illustration of the difference between an after-image being experienced in the external world like perceptions, and I will assume that most people's experience of visual imagery is that it is not experienced in the same way. Hence I will take it as a fairly well accepted general rule that visual images are experienced as non-external. I will also refrain from referring to 'internal imaginary space' from here on and simply suggest more broadly that visual imagery is just **not** experienced externally and leave it open as to how this is then developed further. I will return to discuss this in more detail in Section 7.3.2 and I will use this Rubiks Cube Problem to illustrate other points during the thesis.

1.2.2 –The second question - visual imagery in imagination and memory?

In reply to the **second** question, as described above, it is common to imaginatively go through each room in your house and count off the windows in each room and then add them up (e.g. there are seven in my house). This certainly seems to involve some kind of vision-like phenomenal state that is distinct from what is present in your current surroundings. Specifically this would mean imagining every room and perhaps checking each wall, or at least visualizing the window in each room. Perhaps this can be seen as mentally simulating walking through your house. However at this point it becomes unclear whether you are really imagining this or just remembering it, and this is an example of the ambiguity between imagination and memory. Generally I would distinguish phenomenal memory from phenomenal imagination by suggesting memory is of a past factual situation. Perhaps you could call it spatio-temporal memory, where the temporal aspect is restricted to events or episodes *in the past,* which did actually occur to you.

To illustrate this further, consider another situation where the distinction between memory and imagination gets blurred. For example, remembering a dinner party viewed from above the table, even though at no point during the diner party were you hanging from the chandelier (at least you hope not!). This is a factual situation, which you were present at, but is recalled from a novel viewpoint, which you were never located at. Hence it could possibly be seen as an act of imagination based on remembered facts from the past. So perhaps it represents a combination of memory and imagination and it may be hard to get a clear definition of what counts as each in all cases[11]. In what follows I do not propose to untangle memory from imagination, but only suggest there are definite cases of each with some grey areas in between. For the main purpose of this thesis we can leave the border between the two blurry, because I will be concerned with describing how visual imagery can be seen as an offline perceptual process, and this can explain both memory and imagination. And indeed I will propose a detailed method by which they interact under the Kosslyn Model based on accessing an Associative Memory Mechanism.

1.2.3 - After-images

In terms of after-images, again a good way of getting a feel of these is to just experience them yourself. However in this case I think it is much more important to *actually* do so if you haven't before, because I will use this as part of my argument to suggest that

[11] For examples of this overlap see: Schacter et al (2007), Buckner and Carroll (2007) and Hassabis et al (2008) and also a New Scientist article entitled "Future Recall" (24th March 2007).

externally experienced phenomenal character can be generated by the brain in at least some cases. So try the two below by looking at them for one minute and then looking at a white sheet or a light coloured wall. Also try moving it around by looking at light coloured surfaces at different distances from you and observe the phenomenology:

Figure1.3 – Union Jack and Santa Claus after-images
(see text above and below for an explanation)

What you should see, if it's working properly, is a faded version of the normal colour image of each image experienced as if it is 'out there' in the external world and this can be moved from surface to surface depending where you look. You should note how different this is from the phenomenology of visual imagery, which is not experienced as in the external world and this should give you a clear idea of what I mean by this distinction[12].

Unfortunately the full scientific explanation of this phenomena cannot be given here, but the commonly accepted basic idea is that the receptors in your eyes gets saturated by the colours in the *actual image*, and this causes a neuronal rebound so that for a while afterwards you 'see' the opposite colour in terms of colour processing in the eye (e.g. yellow becomes blue). I take this to mean that this is the equivalent of the eye sending signals to the visual cortex that there is actually an object out there that has these opposite

[12] It is worth noting that I find after-images much easier to generate from looking at them on a computer screen. For a good collection of these and other fascinating illusions go to: http://www.moillusions.com [accessed Sept 2009]

colours[13]. I will introduce a third, slightly more involved, after-image to illustrate some points in Chapter 5. For now this section was aimed at illustrating the difference between two vision-like experiences that are experienced obviously externally (the after-image) and not externally (visual imagery) and to give you some first person experiences of visual imagery to refer to throughout the thesis.

1.3 –The Kosslyn Model and the Tye-like Approach

As mentioned briefly earlier I will defend an approach to mental imagery and STIm that I refer to (following Bartolomeo 2002) as the 'Kosslyn Model'. This theory is a leader in its field and has been developed over the last three decades by several researchers under the main direction of Stephen Kosslyn. Hence I use the term 'Kosslyn Model' for ease of reference only, and recognise that this term refers to the lead researcher in what is an extensive temporal and geographic collaborative effort. I make extensive use of the latest version of this theory as expounded in the book: 'The Case for Mental Imagery', co-authored by Stephen Kosslyn, William Thompson and Giorgio Ganis (2006). There are two key themes in that book that will be crucial to my analysis.

The **first** main theme to be defended is a *depictivist* approach that suggests mental images are processed using hybrid depictive representations. This means I will defend the Kosslyn Model against its main rival, which is the *propositional,* or descriptivist approach (Pylyshyn 2003). I favour the use of the term propositional to refer to this latter position, because it is more easily distinguished from depictivism than descriptivism is (see what I mean!). Hopefully this will make the reading a bit easier and henceforth take propositionalism to be synonymous with descriptivism, and note that I will avoid using the

[13] But see Pautz (2006: 212) for a good explanation of why they take the colours they do based on an 'opponent processing theory'.

latter term. The main claim of depictivism is that visual information is at least partially processed in a depictive manner and this is crucial to the way it represents. That is, it uses space to represent space in a *similar* way that a 2D photograph depicts a real 3D event (I will explain this in detail in Chapter 3). The propositionalist will deny this and argue that any apparent depictive qualities of the representations are a merely contingent feature of the way perception and imagery are processed. I will argue in Chapter 3 that the latest version of the Kosslyn Model seems to deal with most of the latest propositionalist objections, and hence remains very much a viable and useful approach in my opinion. In addition its extensive explanatory power and rich detail make it a favourable one to adopt once these considerations have been dealt with.

The **second** main theme is that visual imagery and perception crucially rely on the visual buffer for processing. The nature of the visual buffer will be explored in detail on Chapters 3 & 4, but a brief description will be given here. Here is an example of how the Kosslyn Model defines it:

> '…at least thirty-two distinct areas of [visual] cortex are involved in visual perception in the monkey brain,… and probably more in humans. Some of these areas (about half in the monkey brain) are topographically organised…damage to discrete portions of these areas produce scotomas – blind spots – that are localized in space according to where the damage occurs… we group topographical organized areas in the occipital lobe into a single functional structure, which we call the *visual buffer*.' Kosslyn et al (2006: 135-136)

> '…mental images are not like points in an array in a computer in at least one fundamental respect: each neuron in the visual area does not simply register the presence or absence of light. Rather, the neurons also code for specific properties such as the orientation of the line segments, hue, and binocular disparity (which is a cue for depth). Thus, although the representation has a depictive component… it is in fact a hybrid representation. Each point is interpreted in part in terms of its role in the depiction, but also in terms of the additional information it codes abstractly.' Kosslyn et al (2006: 18-19)

From the first quote we can see that the visual buffer consists of multiple functionally linked topographically organised areas in the visual cortex[14], each of which process different aspects of our visual experience e.g. colour and depth. The topographic areas are said to be made up of individual cells that together form a visual array, which depictively represents certain aspects of the visual scene in front of you (this is illustrated in the diagram below). However since these cells are spread out on a relatively flat area of the brain, they contain other information within them, which represents things like depth. This is what makes them hybrid depictive representations: they depictively represent 2D space using physical space across the array, but also each cell can contain more abstract contents, and hence may also contain propositional elements. I will develop a method for annotating this in more detail in Chapters 3 & 4, but to illustrate the role of the visual buffer further let's look at what the evidence from scotomas tells us.

Kosslyn et al (2006: 15) describe evidence that shows that damage to certain areas of the visual buffer correspond to loss of detail in corresponding parts of the visual field: i.e. they produce a blind spot or scotoma. And further this is linked topographically because: 'crucially, the closer two damaged regions of the topographically organized visual cortex are, the closer in the visual field the corresponding scotomas will be'. This is illustrated in the diagram below, where I have split the visual buffer (on the left) into cells and the depiction of the visual field (on the right) into pixels:

[14] See Appendix B for a diagram showing the approximate locations of the main areas of cortex and the dorsal and ventral visual processing streams in the brain.

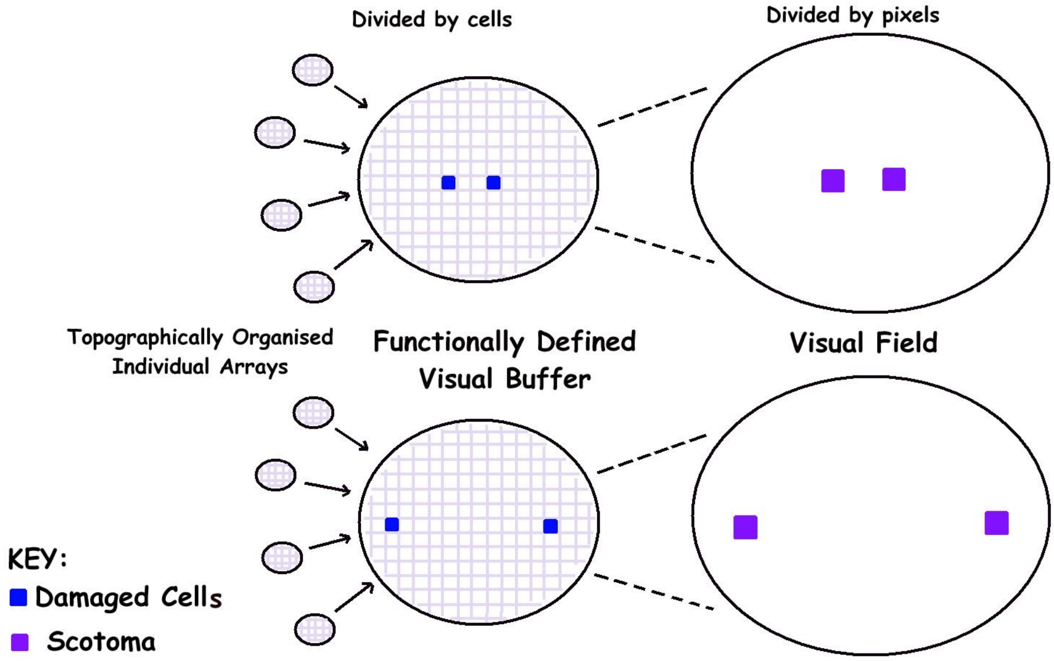


Figure 1.4 – Schematic of the visual buffer array on the left (cells), which is related to visual field representation on the right (pixels). The two diagrams show differently spaced damage in the visual buffer leading to correspondingly spaced scotomas in the visual field.

Hence the main aim of defending the Kosslyn Model here is to look at what each of the cells in the visual buffer contain and how they are filled-in and manipulated. We can also look at the relationship between *cells* in the visual buffer and how they relate to '*pixels'* in the experienced visual field. This gives us a potential correlation between the underlying cognitive processing and the eventually experienced visual phenomenal character, and this theme will be developed further as we progress. I use the term 'pixel' here as a loose analogy with TV-screen pixels and to refer to a very small area of the visual field. But it is important to note that I do not use it to suggest that the visual field is actually divisible in this way; it may be, but this is not a claim I will defend here. Hence this term is not meant to be taken literally as any claim for how the visual field is actually divided up and just refers to a small portion of your visual field.

It should also be noted that the term ‘array cells’ might also be slightly misleading, since each cell is actually composed of millions of neurons that are massively inter-connected and processed in a parallel-distributed fashion. The term ‘array cell’ is a functional term for a group of physically adjacent neurons that behave in a way that can be characterised, and more easily described, by referring to it in an analogous way to the cells in a computer array. Hence the diagram above is highly schematic and I am not suggesting a one-to-one relationship between cells and pixels. The brain is probably much more messy than that, but perhaps we can still talk functionally about certain cells relating to certain areas of the visual field. It is *crucially* important to bear this in mind during this thesis, because at times it may seem like I am over simplifying things. However this is mainly because this is required for the exposition to run smoothly and be aware that I am conscious of the actual complexity of the matter at all times. Saying that, I think the general points I make about imagination-as-simulation will be robust enough to endure a more complex empirical treatment and will remain useful as an approach that can be tested further. I think aligning myself with the Kosslyn Model also supports that prediction and much of Chapter 3 is aimed at explaining and justifying these claims.

In Chapter 4 I will defend the Kosslyn Model against alleged counterexamples to the necessary involvement of the visual buffer in visual imagery. These are purported examples where visual imagery persists even if the visual buffer may not be used or accessible. The main challenge will come from subjects who are cortically blind and have extensive damage to the visual cortex, but yet still claim to have vivid imagery. We will also look at the evidence from the Neglect literature that also shows apparent double dissociations between perception and visual imagery. These terms will be explained in Chapter 4 and this will also give us a full account of the Kosslyn Model that describes how the spatial layout of visual imagery is generated and maintained.

I think this approach is very close to the underlying structure of Tye's (1995) representational PANIC theory. And indeed many of the concessions and adaptations that the modern Kosslyn Model has made are in response to earlier criticisms by Tye (1991). Hence another theme of this thesis is to adopt a very Tye-like approach to at least the processes underpinning perception and imagery. That is even though I will disagree with him on how he eventually explains the generation of phenomenal character (see next sub-section), I do find his underlying 'symbolic array' approach very appealing and still currently relevant. Consider this from Tye to illustrate this point:

> 'There is strong evidence that images and visual percepts share a medium that has been called the "visual buffer"… For visual percepts and afterimages, the visual buffer is normally filled by processes that operate on information contained in the light striking the eyes. For mental images (other than afterimages), the visual buffer is filled by generational processes that act on information stored in long-term memory about the appearance of objects and their spatial structure…. Images and percepts, I have argued elsewhere, are *interpreted, symbol filled patterns of cells in the visual buffer* … I maintain that bodily sensations generally, perceptual experiences, and imagistic experiences all have their contents encoded in arrays or matrices functionally like the sort I have described.' Tye (1995: 122: Box 4.7 – footnotes removed, emphasis mine)

I will discuss what Tye means by 'interpreted symbol filled patterns in the visual buffer' in Chapter 6 and argue that perception and imagery can occur even in un-interpreted arrays; that is as long as certain other qualifications are met. I will also make some distinctions between different kinds of possible meanings of 'interpretation' and propose a theory that suggests that willed images might need to be *labelled* (or interpreted in one sense) to be useful. Conversely I will also argue that phenomenal images can occur even if they aren't examined or conceptualised (or interpreted in another sense). That is, they may go unnoticed because attention is elsewhere, or they may be very abstract, or non-conceptual, or vague, and hence be fairly un-interpretable. They may also occur spontaneously and automatically and hence their presence and contents may be surprising and sometimes unwelcome (e.g. distressing flashbacks).

Despite that, I will argue that if you are using visual images to consciously work something out, then they will need to be attended to, and have an appended context (or spatio-temporal label) and be associated with an act of will that is aimed at reaching a desired goal. The Rubiks Cube Problem described above is an example of this and I will return to elaborate on this further in Chapter 6. The net effect is that eventually we will have a fairly comprehensive explanation of how visual imagery occurs and how it can come to be *labelled* or *interpreted* as representing a non-current spatio-temporal setting. Hence, after certain intermediary steps are completed, this hopefully leads us to a fairly comprehensive theory on the nature and function of STIm. For now we have a good introduction to the first half of the thesis and an indication of why I think it is relevant to the overall aims of the project.

1.4 – STIm and Vision: Selective vs. Generative Approaches to Perception

In the second half of the thesis (Chapter 5-7) we will start to look at what taking an online visual mechanisms offline to form phenomenal visual image might involve. This will take the form of discussing what certain perceptual theories say about the nature of online visual perception and investigating the ramifications of this for their offline equivalents. One of the key distinctions I will utilise in the thesis is that between perceptual theories that suggest all phenomenal character supervenes locally on the brain and those that suggest it supervenes more widely and *constitutively* depends in some way on the object itself. A good way of illustrating this is to use a distinction made by Howard Robinson between S*elective* and G*enerative* Approaches to perception. A **Selective Approach** is best associated with relational[15] theories like Naive Realism and suggests that the causal neuronal process 'does not generate a content but puts one in touch with the stimulus that is already out there' (1994: 70). That is the phenomenal character widely supervenes on the object, the brain and all the causal processes in between and properties you are aware of (e.g. colour) *can* belong to the object itself. A **Generative Approach** on the other hand suggests that the neuronal areas involved in perception, not only detect and process the stimulus, but they also generate the phenomenal character too. That is, the phenomenal character itself supervenes only locally on the brain, but the classification of the kind of phenomenal episode it actually is, also supervenes on how it is caused: i.e. if it is caused by an external object then it gets classed as a perception or illusion; if not then it is a hallucination or a visual image. Classic examples of a Generative Approach are most indirect realist theories such as Sense-Data Theory and certain types of Qualia Theories. This distinction and its implications will be explained in more detail in the introduction to Part II of the thesis, but I will give a brief summary of that here too.

So why is this distinction useful in this context? This is because I will argue that if some form of *Generative Approach* to perception is favoured, then it seems we already have potential mechanisms in the brain for also generating phenomenal imagery. If a *Selective Approach* of perception is favoured then it seems no such mechanism is readily available, and it is possible that we may need a separate one for locally generated phenomenal character. **So my main claim here is this**: that perhaps by looking at things in this way, we would predict much more overlap between imagistic and perceptual mechanisms in a Generative Approach, because they might use the same underlying mechanisms and be generated in similar, yet slightly different, ways. Note that the claim in not that there is a complete overlap, because we still have to explain the differences between phenomenal

[15] I use the term 'relational' here to distinguish it from 'representational' theories as per Crane (2006); this will be explained further at the introduction to Part II of the thesis.

character like perception, that is experienced as if it is external, and visual imagery which is experienced non-externally. So although they may both be generated locally by the brain, they are somehow represented slightly differently and hence experienced differently. Thus under an imagination-as-simulation approach we would expect a fairly extensive overlap in certain crucial mechanisms (e.g. visual buffer), but also some differences in how images are experienced phenomenologically speaking. This will be explained in detail in Chapter 7.

On the other hand if you adopt a *Selective Approach*, then it appears possible that all your brain does in perception, is somehow make you aware of the properties (e.g. colours) that external objects have independently of your mind. In which case the simulationist will predict that there is much less overlap between perception and imagery, because in imagery the brain perhaps also has to generate the phenomenal character, since you obviously can't just access it from the external world (if you could there would probably be no point in imagining it!). Hence it is proposed that by looking at different perceptual theories on either side of this divide, the simulationist might have to make different predictions about the nature of the relation between perception and imagery. Analysing this is one of the core aims of the second half of this thesis. Figure 1.5 illustrates one way that this difference could be functionally depicted:

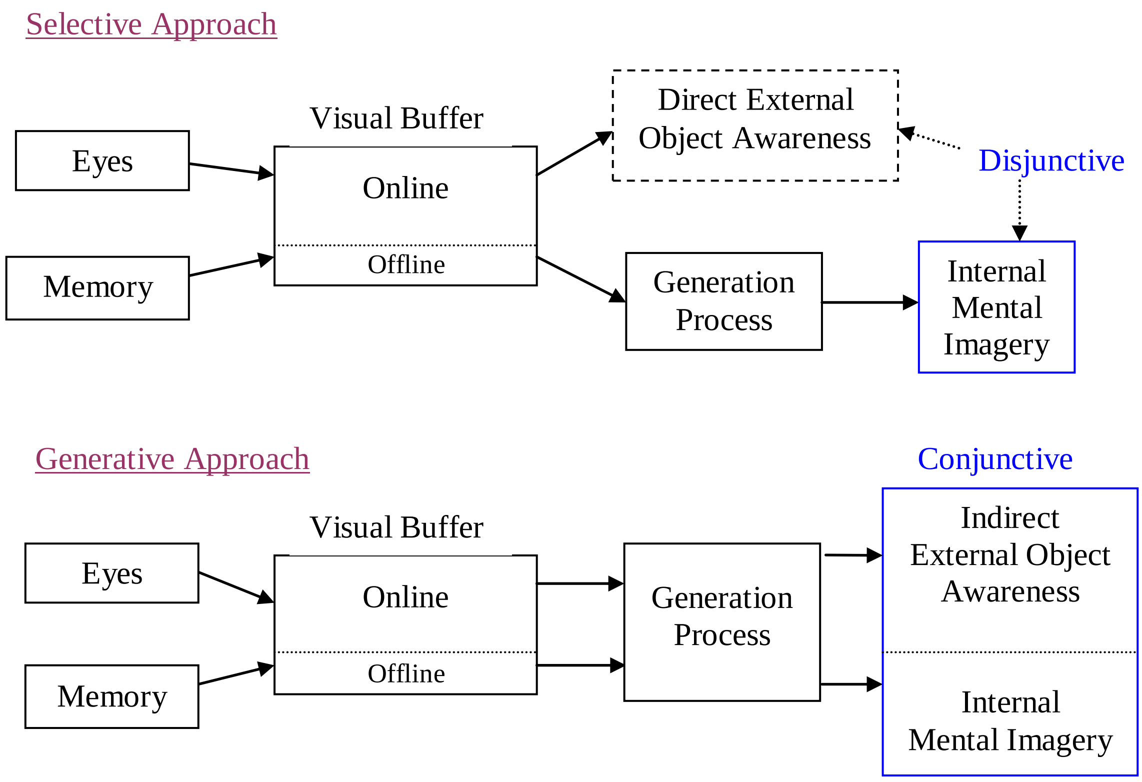


Figure 1.5 – Functional depiction of a Selective Approach (above) and Generative Approach (below) to perception – showing the extra generation process possibly required for imagery in Selective Approaches.

So in order to analyse how close the relationship between perception and imagery is, I propose that it would be useful to see what different theories of perception might say about the imagination-as-simulation debate. In terms of a Selective Approach I intend to use modern Naïve Realism[16] as my paradigmatic example and this categorization is supported by the fact that at least one current Naïve Realist (i.e. Fish 2009: 137), specifically refers to it as such. Hence I will analyse a modern expositions of this theory[17] in Chapter 5 in order to examine its potential consequences for the imagination-as-offline-simulation debate.

[16] The 'modern' qualification distinguishes it from old style classical naïve realism which is described by Le Morvan (2004: 222) as: 'a strong form of Direct Realism, [which] claims that perceived objects or events always appear exactly as they are.' Which obviously falls prey to the arguments from illusion and hallucination.

[17] C.f. Martin (2004, 2006), Fish (2008, 2009) and Kennedy (2007, 2009)

I will also access the burgeoning Disjunctivism debate where necessary[18], as this is crucially linked to understanding Naïve Realism. This is because, according to some: '…the prime reason for endorsing Disjunctivism is to block the rejection of a view of perception I'll label *Naïve Realism*' Martin (2004: 38). Disjunctivism is distinguished by its key tactic of separating off vision and vision-like states with a disjunct like the following: EITHER you are in a visual state of a certain kind (e.g. veridical perception); OR you are in a non-veridical vision-like state of a significantly different kind (e.g. hallucination). The idea here is to separate veridical perception off as a different kind of thing from phenomenal hallucinations, even though they may superficially look like the same thing. This helps protect Naïve Realism from the Argument from Hallucination used against Selective Approaches and this will be discussed further in Section 5.1. Related to that, following Johnston (2004), I will refer to a theory that denies Disjunctivism, as a Conjunctivist Theory and you can see these labels being used in Figure 1.5 above. Conjunctivism would accept that there is a significant common overlap in the kind of things perception and hallucination are, in that they may both be phenomenal experiences that are locally generated by the brain (e.g. both are kinds of sense-data or qualia)[19]. Again I think Figure 1.5 illustrates this difference between these approaches quite well and this will also be discussed further in the introduction to Part II of the thesis.

This approach therefore connects this discussion to those about the amount of 'common core' or 'common factor' between perception and hallucination[20]. Disjunctivists deny any such commonality and it seems that this can be potentially extended to my analysis on the

---

[18] C.f. Haddock and Macpherson (2008) and Gendler and Hawthorne (2006).

[19] For example Johnston (2004: 114) says: 'Any Sense Datum Theory that treats sense data as the only objects of immediate awareness… is obviously a version of a Conjunctive Analysis. But other models… such as Adverbial Theory and various Intentionalist accounts of visual experience, can also take a Conjunctive form'.

[20] C.f. Crane (2005) for a good introduction to the use of these terms.

common overlap between perception and imagination. Consider this quote by Martin from a compilation on 'Time and Memory' to illustrate this point:

> 'The moral to emphasize at this point is just that the assumption at work in so much philosophical discussion of memory is that the phenomenological or the sensational must be a common element of experience and imagery can and should be denied. We should not accept that there is a common core of sensation which is neutral between perception and imagination or memory of such perceptual encounter. Imagination and memory relate to perception not through replicating the sensational or imagistic component of perception, but through being a form of representing such experiential encounter with the world.' Martin (2001: 273-274)

Hence we can see that there is an explicit denial of a common core between perception on one side and imagination and memory on the other. It turns out that this denial depends on some fairly heavy implicit Naïve Realist and Disjunctivist assumptions, which would be denied, for example, by a Representationalist and Conjunctivist like myself[21]. I will investigate what this denial of a common factor means for the imagination-as-simulation approach in Chapter 5. There I will also reject Naïve Realism by offering an updated version of the Time-lag Argument, which I call the T*ime-spread Argument*. This is aimed a justifying why I do not favour any version of a Selective Approach to perception.

I will then explain and analyse Tye's (1995, 2000) representational PANIC theory in a bit more detail in Chapter 6. This will start by explaining how his underlying array based symbol manipulation theory still seems very compatible with the updated Kosslyn Model. It will then move on to look at how this underlying structure leads to phenomenal consciousness in perception and imagery. Tye's PANIC theory is often treated as a common factor theory in the literature, as there seems to be a significant overlap between perceptual and imagistic states as illustrated in the quote from him above. However I will

[21] In fact it was uncovering these implicit assumptions (and rejecting them) that inspired much of the way that my PhD has turned out.

criticise his position for being overly representationally reductive[22] and try to assess whether his theory is best interpreted as a Selective or Generative Approach to perception. I will argue that I think it is best seen as a Selective Approach and therefore I will also reject it by applying the Time-spread Argument to it.

In the end I will eventually defend and adopt a representationally non-reductive Generative Approach to perception in Chapter 7. I take this to be a general term that lumps together theories that suggest phenomenal character supervenes locally and is also generally representational in nature. It however denies that phenomenal character can be reduced to representational content and hence some further feature is required (e.g. sense-data or qualia). I will also assume a Generative Approach will usually accept that the following Brain in a Vat (BIV) intuition:

> 'At some point your brain is stopped so that you cease to experience anything. Your brain is then removed and placed in a VAT at which point it is wired up so that it receives exactly the same input. Your brain is then restarted and you cannot tell the difference.' – BIV intuition

I also think this is compatible with the claim that a microphysical duplicate of your brain will be having the same experiences as you and this is because the brain generates the same phenomenal experiences in both cases. I think examples of this kind of approach potentially include Sense-Data Theories, certain forms of Adverbialism, certain Qualia

[22] Tye's (1995) PANIC theory is classed as representationally reductive because it claims that phenomenal character can be reduced to representational content without remainder, and hence there is no need to invoke extra entities like qualia (see Chapter 6 for more on this). It is an odd feature of the timing of this thesis that it will be completed just after Tye (2009) has published a more updated version of his current position. In this new work he specifically rejects approaches based on 'phenomenal concepts', since he now thinks these are indefensible. Since the appeal to phenomenal concepts was a crucial part of his defense of the *representationally reductive* nature of his PANIC theory, it seems this may entail a wholesale rejection of his previous position. Unfortunately he is not specific on how extensive this rejection or modification may be, so this is a point that is still in need of further clarification.

Theories and the newly developing set of theories that fall under the banner Phenomenal Intentionality[23].

However since all I crucially need to do is to defend any form of a Generative Approach, because I think this is the crucial thing to ascertain in terms of the imagination-as-simulation debate, I will try to stay as general as possible in this area. In other words I will try to keep it open as to which specific Generative Approach to perception I think is the best for developing this account further. Although I will say specific things about specific versions of these theories as we go along, I will try to stay general as possible in terms of committing myself to one particular generative theory or another. This will be developed and explained further in Chapter 7, where I will conclude that there is a significant common overlap between imagination and perception in that their nature is best captured by referring to them as: *locally generated visual style phenomenal representational mental states*, with relatively minor differences in how they are caused and experienced.

I will finish by suggesting a method by which these visual images become labelled with a different spatio-temporal contexts, and how these may function in our wider economy, and hence this will indicate what I think the nature and function of STIm is under an imagination-as-simulation approach; which is what the title of the thesis asks. So answering this question is the main aim of the thesis and the summary above gives you an initial idea of how I plan to go about it. For now we can begin this journey by getting a more detailed explanation by what I mean by imagination-as-simulation in the next chapter and by looking at other researchers who may be appealing to similar things. This will give us a platform to work from for the rest of the thesis and also put my discussion on visual

[23] C.f. Robinson (1994) for a modern sense-data theory; and Wright (2008) for a collection of qualia theories; and Horgan and Tienson (2002), Horgan and Kriegel (2008), Kriegel (2007a) for papers on Phenomenal Intentionality. And note that Kriegel (2007a) also adopts a form of Adverbialism.

imagery into the wider context of discussions of other forms of mental imagery: e.g. motor imagery.

Part I: STIm and the Mechanisms Supporting Visual Imagery: a defence of the Kosslyn Model under an imagination-as-simulation approach

2 - STIm as an Offline Simulation Process

"Evolution is cleverer than you are." - Orgel's Second Rule,
attributed to evolutionary biologist Leslie Orgel (1927 – 2007)

In this chapter I will discuss in detail what I mean by an *offline simulation process* and relate this to my investigation of STIm. I will do this by looking at some generic examples of offline simulation processes and explain how some other researchers utilise this notion in their work. This will involve discussing situations where offline simulation processes are utilised to explain motor imagery. This analysis will give us some key features that any offline simulation process should possess and this can then be used as a platform to argue that offline simulation is the main process that sub-serves visual imagery in what follows. Thus the main aim of this initial analysis is to provide some useful distinctions that can then be applied to the following chapters on the relation between visual imagery and perception.

Therefore in this chapter I will aim do the following four main things. **Firstly**, I will introduce some cautionary notes about this style of approach and highlight things that we should be aware of throughout the course of the thesis. **Secondly**, I will explain what I mean by a simulation process and introduce a few varieties of these. I will then explain how offline simulation is a specific version of a simulation process and outline some features this kind of process may have. **Thirdly** I will relate this to an *emulation* system in the motor imagery literature (Grush 2004), which has certain features in common with simulation, but also certain crucial features that are different. This will also allow us to investigate in Chapter 4, whether there is such a thing as a visual emulator that could be harnessed in visual imagery. **Finally** I will link this with work done by Susan Hurley (2008) on her Shared Circuits Model, where she combines emulation and simulation to underpin a mental feature she calls *deliberation*. I take deliberation to be roughly equivalent to strategy testing and hence this aligns her style of approach fairly closely with mine in certain respects. This is because we both want to explain a capacity for planning based on more simple emulation and simulation processes. The combined effect of this chapter is aimed at giving us a clearer idea of what counts as an offline simulation and provide further insight into the nature and function of STIm.

Before we continue I should make it clear that, although I will mainly be focusing on concerns within the field of Simulation Theory during this thesis, I do *in general* adopt what is called a Hybrid Theory. This is an approach that accepts that a combination of Theory Theory and Simulation Theory is required to explain our full range of mental capacities[24]. For instance, in mindreading we may need to form theories about other people's mental states as well as simulate what they might be thinking. However since Theory Theory is not quite as relevant to the mental imagery debate, this explains my exclusive focus on simulation theories in this context.

2.1 – Cautionary notes about simulation approaches

Before we continue I will briefly go over three cautionary notes about using this kind of approach, which I will to heed during the thesis.

[24] See Carruthers & Smith (1996) and Davies & Stone (1995) for general collections on this theme and Stich and Nichols (2002) for the use of the term 'hybrid'.

2.1.1 – Personal vs. Sub-personal Levels Caution

The first thing to be aware of is that simulation theories can be investigated at the sub-personal or personal level and the two aren't *necessarily* connected. I shall define the *personal* level as any mental phenomenon that is accessible to us in a first person way. These include mental features that we are currently conscious of and those we can potentially become conscious of which may currently only inhabit our sub-consciousness: e.g. the sound and feel of our breathing. Alternatively, *sub-personal processes* as those that aren't accessible to us as a person. This can include neural processes or functional groupings of neuronal processes described at the level of underlying mechanism. These are ones that we can never be conscious of and are more to do with the basic workings of our brain: e.g. breathing regulation. Although this exact nature of this distinction is currently debated[25], in terms of this thesis it is fairly clearly demarcated. The distinction we will be concerned with here is between phenomenal episodes that we consciously experience at the personal level (e.g. phenomenal imagery) and the functional neuronal mechanisms that may sub-serve these processes at the sub-personal level (e.g. the visual buffer).

This warning was first put forward by Heal (1998), where she makes an important distinction within the simulation literature that deals with mindreading. She refers to personal level mindreading as co-cognition, where all she means is the shifting of our perspective to simulate another point of view: e.g. ourselves in another time or place or ourselves in the position of another person. This can be seen as a separate issue as to whether this mental ability is based on a simulation process at a lower sub-personal level. I will describe three main versions of simulation at this lower level in the next section to illustrate what this could mean.

---

[25] C.f. Bermudez (2000), Hornsby (2000) and Dennett (1968).

The point of introducing this caution here is to highlight Heal's main concern that failure to describe things at one level doesn't necessarily indicate failure at the other level. Hence if you are not convinced of arguments about sub-personal simulation mechanisms in the brain, this can potentially be separated from discussions about simulating other perspectives at the personal level (i.e. co-cognition). However having highlighted this concern, this thesis will try and make some links between these two levels of description. But the reader should bear in mind that it is possible that these two agenda's could be treated separately if so desired[26]. The relevance of this distinction to this thesis is as follows. The first half of the thesis will mainly focus on sub-personal offline simulation processes that might support visual imagery (Chapters 2-4). And the second half of the thesis will look at how these may relate to our personal level phenomenal experience (Chapters 5-7).

2.1.2 – Interlevel Isomorphism Assumptions Caution

In what follows we should also bear in mind the following warning from Susan Hurley. Hurley (2008: abstract) suggests that her theory: 'is cast at a middle, functional level of description, that is, between the level of neural implementation and the level of conscious perceptions and intentional actions'. This is because in a lot of cases she is looking at subpersonal mechanisms that *make possible* or *enable* personal level experiences. In what follows I sometimes refer to this as these subpersonal processes *underpinning* the personal level ones.

Hurley is careful to respect this personal/subpersonal distinction and avoid 'interlevel isomorphism assumptions'. That is descriptions of functions at the subpersonal level need not share structure with personal level descriptions. Obviously it is a matter of debate as to

[26] It is also possible that Goldman (2006: Chapter 6&7) makes a similar distinction between high and low level simulation in terms of mindreading.

how much isomorphism there is between levels and this can be addressed by looking at how closely specific personal level capacities are controlled by, and match up with, these underpinning subpersonal mechanisms. The cells to pixel relationship briefly introduced in the last chapter is perhaps an example where there may be a close link, but we must be wary of over simplifying things by invoking a one-to-one relationship and just *assuming* interlevel isomorphism. The brain to mind relationship is a complex one and we may only be able to make some tentative suggestions about it here. This issue will be returned to and discussed in more detail in Section 7.2.

2.1.3 – Boxological Flow Diagram Caution

Throughout the thesis I will continue to use 'boxological diagrams', similar to the one I used in Figure1.5 in the last chapter. However there are a few it is worth making here about what they are supposed to represent and their depictive limitations. Stich and Nichols (2003) make some useful points in this area[27]. They utilise some functional box diagrams in their expositions in order to illustrate a: 'functionally characterized processing mechanism or a functionally characterised set of mental states' [pg. 11]. Hence a lot of their book is focused on introducing and defining certain mental functions that they think are crucially involved in reasoning, mindreading and pretence. And they think showing how they interact is sometimes usefully represented in a boxological flow diagram. They justify their 'functional analysis' by suggesting that it is the first step to understanding how the mind works and cite a certain amount of philosophical pedigree that supports this approach (e.g. Fodor 1968, Dennett 1978, Lycan 1981, 1988).

[27] Stich and Nichols (2003) also introduce and defend a 'Possible Worlds Box' mechanism that is involved in pretence, planning and mindreading. This is invoked to allow the subject to entertain imaginary *beliefs* in order to consider alternative possibilities and guide our actions. I would suggest that this is very close to my idea of what STIm is crucially involved with in strategy testing but with some crucial differences. I have analysed the Possible Worlds Box as a form of imaginative capacity in my MA thesis (King 2005c), but I will unfortunately not be able deal with it any further in what follows. See also Heal (2003, 2005).

However they warn that this way of depicting processes is not to be confused with the idea that these functionally specific mechanisms are localised in the brain. The prediction of a *functionally* isolated system says nothing about how they are *physically* realised. They could be realised very locally or distributed throughout the brain, as long as they serve to provide that function it doesn't really matter how or where they are located. But obviously this is where brain imaging and brain lesion studies can help us identify certain areas that roughly correspond to certain functional mechanisms in the brain. Perhaps a potentially limiting overriding constraint is that areas that often communicate with each other are more efficiently physically located near each other where possible. Its possible that the *visual buffer* is a good example of different topographical areas located reasonably locally in the visual cortex, that might best be treated together functionally in the Kosslyn Model; more on this in the next two chapters.

Commenting on the use of these 'boxological diagrams', Heal (1996: 51) suggests we must be wary of the way they oversimplify the interactions in the mind. For instance, there are vast amounts of inputs and outputs going in and out of each actual mental mechanism, and these may subtly affect the experiences we have and the decisions we make. Depicting this process with a few one-way arrows that only interact with a single box gives the impression that only these interactions are possible or allowed. In reality the interactions may be much more complex, bi-directional and distributed, and indeed the types of knowledge, memory and learning we bring to bear on a decision may be vast and subtly interrelated. Hence Heal suggests we need to bear in mind the 'epistemic holism' of how the mind works, which is a feature that is possibly lost, or severely under represented, in these kind of diagrams. These points will become crucial when I defend the Kosslyn Model in the next few chapters. The take home lesson is that one shouldn't take these schematic diagrams too seriously, since their only purpose is to help illustrate certain points in the text in a simplified and accessible fashion. It is merely hoped that this will aid in the subsequent analysis, rather than be taken as a thorough representation of the complex workings of the human mind. With these cautions in mind we are perhaps now ready to look at different simulation processes in more detail.

2.2 – Offline Simulation

In this section I will outline my definition of 'offline simulation', which I will subsequently adopt for the remainder of the thesis. This will provide a more detailed idea of what I mean by the term so this can then be elaborated on and utilised in the following chapters. Perhaps the best way of explaining this is to give some examples of it and show how it differs from some other main kinds of simulation processes.

2.2.1 – Basic offline simulation

The most simple way of explaining offline simulation is to describe it as a process by which you take an already existant functionally defined main mechanism and feed it mock inputs so as to produce mock outputs, in order to know what the real inputs might have been had the inputs been real. A mock input or output is basically the same as a normal real input except that it is *flagged* or *labelled* as merely imaginary or pretend. This means that it is not generated from a real stimulus or as a response to real current events. This can take various forms and be generated in various ways and examples of these will be given below.

When the main mechanism is working normally, using real inputs and outputs, it is said to be working online. The process of feeding a mental mechanism mock inputs, and subsequently generating merely mock outputs, is what defines it as being taken offline. The merely mock input-to-output relationship should in some way mimic or simulate what you might have occurred had the situation been real. For example, if processing information from items in our environment during visual perception is the online function of the visual buffer, then arguably the visual buffer can be fed offline mock inputs (based on memory), in order to generate offline mock outputs as visual images; or in other words it can generate imaginary visual scenes instead of actual visual scenes. These visual images can be said to mimic or simulate what it might be like to see the same events actually occurring. But obviously there will be some phenomenological differences and these will be discussed and accounted for in what follows. Perhaps the offline simulation process can be illustrated in the following generic way:

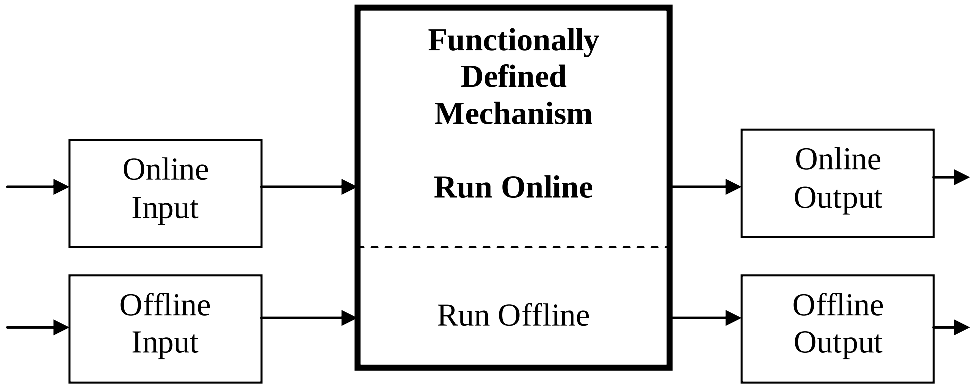


Figure 2.1 – Schematic of basic offline simulation process

The crucial thing in the diagram above is that the online mechanism is re-used in the offline process, but the mock flag attached to the inputs prevent this from being treated normally and this keeps the mock outputs isolated from online action control. This is important because we don't want the self-generated offline mock signals to be confused with real online environmental input. For example, should we decide that we need to visually imagine a tiger, it is important that the visual image does not cause us to panic and run away and potentially hurt ourselves unnecessarily. Conversely it is important that an online perception of a real tiger does lead to urgent evasive action and is not confused for a relatively harmless visual image of a tiger. Obviously using the same mechanism for two different purposes runs this risk of confusing the two, but generally it is assumed that this separation could be hard wired in such a way, that failure to distinguish the two rarely occurs. It is possible that certain forms of hallucination and other mental problems can be explained by a situation where this process goes wrong and we will touch on this more in what follows.

As a way of illustrating this process perhaps we can look at how this would work in a generic formulation of an offline decision-making process, of the kind that is sometimes invoked by simulationists in the mindreading literature. Here it is suggested that we have a decision-making mechanism that usually works online to guide decisions based on the current intentional mental states of the subject: e.g. their beliefs and desires. That is, it takes the relevant beliefs and desires of the individual as an *input* and processes these along with other information (e.g. perceptions, memories), to *output* a decision, in order to guide what the best next course action is likely to be. In an episode of imaginative decision-making this would only generate pretend beliefs and desires, which can then be used as mock inputs into the same system. These would be flagged or labelled as such to prevent them from being acted on as if they were our real current beliefs and desires. The decision-making mechanism could process these nearly as normal and the mock label would be transferred to resulting mock decision output. This can seen as the equivalent to the decision you would have made had the original beliefs and desires been real and not just imaginary. This mock decision could then possibly be assessed by comparing it with other pretend decisions, which in turn have been generated from a different set of pretend beliefs and desires, and the most beneficial one identified and eventually actually carried out. Note that I am not attempting to defend this approach here in any detail. I merely introduce it as an illustration of one way that the features of offline simulation process could be realised in another field of enquiry. This is aimed to help indicate how this process might work in the arena of visual imagery.

2.2.2 - Varieties of simulation

Another way of illustrating what is special about offline simulation processes is to highlight the differences between it and other varieties of simulation. To that end, I would suggest that offline simulation differs from a *computer simulation*, because in offline simulation an actual real mechanism is somehow involved in the process and in a computer simulation it is not. Taking an online mechanism offline would be difficult to do if you were, for example, modelling the weather, since you cant really feed mock inputs into the weather system. I will call this kind of computer based simulation process a ‘full simulation’, since all parts of the system (e.g. the weather), are modelled in a completely new medium (i.e. binary in a computer). Another kind of simulation process is where you make a scaled model of the original system and test it out using scaled down versions of the conditions you want to test it under. For instance, building a scale model of a bridge and testing it in a wind tunnel in order to see how a real bridge would react to windy conditions. I call this a *scaled simulation*, since not only is the model scaled, but also the prevailing physical conditions have to be scaled down too. Conversely the results would need to be scaled up so as to predict the relevant outcomes in the actual situation.

It may be worth noting that out of the three types of simulation identified above, offline simulation, arguably and broadly speaking, seems to require the least amount of effort, since you are re-using much of the online system already present. The main trick is to work out how to make it work offline and make the inputs and outputs meaningful, and how to keep them isolated from the online system. This potentially adds parsimony to its attributes and this is something that may make it more amenable to an evolutionary process (see later). In what follows then, we will mainly be considering ‘offline simulation’ and ‘full simulation’, and specifically looking at versions of this that might work within human brains. For instance, I will discuss in the next section if a ‘motor emulator’ mechanism is best interpreted as being a full simulation process. I would also suggest that full simulation and offline simulation are the most obvious candidates for potential sub-personal mechanisms in the brain. Conversely scaled simulation seems less clearly appropriate and consequently will not be mentioned again in what follows: i.e. no one would want to suggest we actually use scaled down models of the external world in our brains.

At this point let me offer a working definition of offline simulation:

> **Definition – Offline Simulation**: Where part of an online real system is fed mock inputs, which it processes in a similar way to real inputs, but which only yield mock outputs (in normal cases). This is done in order to provide information about a situation where the inputs and outputs might have been real, which can then usually be utilised for other purposes, such as guiding real online actions.

To that end the main focus of this thesis, will be discussing whether visual imagination can be seen as an offline simulation of the normal online visual process. The hypothesis to test will be to see whether the mechanisms normally used for vision can be fed mock inputs, in order that a visualisation (as a mock output) of a non-presently-occurring scenario is produced. We can then propose that this willed phenomenal experience could then potentially be manipulated and accessed to test out alternative possible scenarios. From this the most appropriate course of action can be chosen, in order to execute this actual action in the present. It remains to be shown how these mock inputs and outputs can be produced and interpreted, and how this 'running offline of the visual system' can occur. But I think this now gives a fairly robust initial definition of offline simulation to use throughout the thesis[28].

---

[28] It is worth noting that offline simulation seems to roughly match up with that referred to as 're-use simulation' in a recent Simulation Theory exchange between Goldman (2008) and Hurley (2008b). It seems that re-use simulation necessarily involves the re-use of the original process in order to simulate something. This discussion stems from a distinction made by Goldman (2006: 36) in a book called 'Simulating Minds', which deals mainly with mindreading. Re-use simulation is contrasted with resemblance simulation because the latter only implies that a second process simulates the first process if it 'duplicates, replicates, or resembles' the original process in some significant respect relative to the task. It seems that resemblance

2.2.3 – Features of Imagination-as-Offline-Simulation

The next stage is to look at what features we would expect from an imaginative mechanism that is based on an offline simulation process. This is the foundation of an imagination-as-offline-simulation approach. One important step in this analysis is to ascertain why this kind of mental mechanism may have developed like this is the first place. Recall that in Chapter 1, I have quoted Currie (1995b) as suggesting the proper function of the imagination is to *strategy test*. Well perhaps we can suggest that once this feature evolved, it seems possible that any increase in its scope and range might have also benefited the organism. That is, the more things it can bring into consideration and the further it extends its ability to plan ahead and learn from the past, then perhaps the better it will do in life. Hence any increase in the range and scope of STIm could have also been selected for once it arose; that is as long as the benefits it yields outweigh the cost of evolving and running it in the first place.

Talking like this seems to parallel some ideas presented by Dennett (1995), who envisages an early evolutionary development stage he calls a 'Popperian Creature'. Here the organism can model the external environment internally and review possible future actions, thus permitting *its hypotheses to die in its stead* and allowing it to choose the best course of action in advance. The suggestion is that the creature has an internal model that is:

> '…an inner something-or-other that is structured in such a way that the surrogate actions it favours are more often than not the very actions the real world would also bless, if they were actually performed. In short, the inner environment, whatever it is, must contain lots of information about the outer environment and its regularities. Nothing else (except magic) could provide preselection worth having.' Dennett (1995: 375)

---

simulation category incorporates all of my varieties of simulation above, but only offline simulation can be identified with re-use simulation.

So not only can the organism internally represent its environment but it can also manipulate this representation in order to simulate different future actions and assess their outcomes. This inner model of external reality could be interpreted as an equivalent way of talking about an early imaginative capacity or a proto-STIm: i.e. the very beginnings of the ability to recreate a different time and/or place in order to strategy test[29]. Given the above, the simulation theorist might prefer to talk of running already existant mechanisms offline, rather than creating an internal model, but the evolutionary benefit of developing something like this is the same in each context. Hence maybe these approaches could just be seen as different ways of explaining the same thing; but perhaps the simulation approach offers a more efficient explanation of the initial way this could have developed.

To support this idea, Currie and Ravenscroft (1997: 177) and Currie (1995a) present evidence that visual imagery and motor imagery are parasitic on the already existent processes of visual perception and motor action. They use this to argue that these imagery tasks can be interpreted as simulation based processes and hence we can see that this method of re-using and running offline already existent processes may have been applied a few times to different imagination tasks. So what perhaps further strengthens our claims for simulation based explanations, apart from its proposed relative simplicity, is that if evolution hits on a good trick to solve an environmental challenge (i.e. offline simulation), it makes more sense to suggest that this same trick is used again in a different domain, than to propose a whole new process or entity to do the same job. They suggest that simulation is one such problem solving trick that has been used again and again by evolution. It is this

[29] I have analysed a plausible story for the early evolution of imagination-as-offline-simulation elsewhere by taking based on taking primitive stimulus-response pairings offline (King 2005a). It should be noted that Dennett was not specifically proposing a simulation theory here, but I think my approach could be one way of developing his ideas. Another researcher who discusses a similar thing in terms of the evolution of 'scenario visualisation' is Robert Arp (2005, 2008). Unfortunately these are also aspects of this field of study that I will not be able to explore further here, but intend to in future. See also Carruthers & Chamberlain (2000).

parsimony, *all other things being equal*, that might support simulation based theories over other theories that bring something new into the world. This is of course only if all the other features of imagery can be accommodated. This thesis is designed to develop this hypothesis by testing and updating further.

A further point to make is related to the relative efficiency of imagination-as-offline-simulation. Currie and Ravenscroft (2002) warn of a potential confusion in this area, where someone might propose that two steps are required in an act of simulation: first pretend inputs are generated; and then the original system being used in the simulation is stopped and then 'taken offline'. This may be seen as a fairly inefficient process that may end up with some strange consequences. For instance, according to this method shouldn't we go temporarily blind as your visual system is taken offline during imagery; or: '…we would always fall in a heap when we mind-read' [pg. 70]. Obviously this is not what happens when we imagine a novel scene in the visual medium or when we take our decision making system offline. So perhaps a better way of describing this process invokes only one step: the feeding of pretend inputs into the real system (e.g. memory traces of cats); that go through a small part of the online system as normal; but only yield pretend outputs (visual images of cats). While this is occurring the main visual system can operate *nearly* as normal, although there will be a slightly extra demand on our attentional abilities and possibly a drop in the information being processes from the external world. So this description potentially fits more closely with the phenomenology of this event and matches what the empirical evidence tells us as mentioned in Chapter 1. How this might occur in visual imagery will be elaborated on in Chapter 4, but for now we can see the point that imagination-as-offline-simulation can possibly not only be motivated by efficiency constraints, but also by potential evolutionary benefits and a good fit with the current empirical evidence. All of which, I think, adequately motivates further research, at the same time as elaborating on other features of this approach.

Although what I have discussed immediately above does appeal to the relative simplicity of imagination-as-offline-simulation approaches, we do need to be aware that this 'taking offline' may also need to be *instigated* and *interpreted* for it to work properly. By *instigation* I mean the kind of thing that prompts the episode of imagination to be initiated in the first place. Perhaps this could be because it is utilised as a way of working out the answer to a question, such as the Rubiks Cube Problem introduced in the last chapter. By *interpretation* I mean the subsequent processes that make use of a mental image. This could involve evaluating the useful information that the mental image is supposed to be delivering and how it relates to the initial task it was designed to help with. It seems that elaborating on these other steps may involve introducing a few more mechanisms into a full explanation of the processes involved in mental imagery to those depicted in Figure 2.1.

It should be noted however that many of the same additional processes would potentially still need to be invoked in other non-offline-simulation based alternative approaches for explaining imagination. It may also be that many of these extra processes could in turn be parasitic on already existing online mechanisms. So while adding these mechanisms may add to the complexity of this explanation, it does not necessarily add to its relative inefficiency when compared with other approaches. We can also make a distinction now between the relative simplicity of explaining how imagination-as-offline-simulation *may have evolved*, as opposed to the complex faculty of reasoning based on imagination *that humans posses today*. The increased complexity in what follows is aimed at providing a structure that can capture the nuances of our intricate modern imaginative capacity. It is also specifically designed to match the intricate nature of the Kosslyn Model to be discussed later.

In terms of what other mechanisms might be needed for full blown modern STIm, Currie (1996: 249) has in the past suggested the following mechanisms might be necessary for offline simulation to run effectively: *a simulation motivator, an input identifier, a feeder mechanism* and *an output detector*. The four components were given as a rough example of how simulation theories could possibly explain the wide variability of disabilities found in autism. This is because different mechanisms could be affected in different ways to potentially explain different autistic deficits[30]. For our purposes here we do not need to go into the details of the application of this for potentially explaining the spectrum of autistic disorders; although this is something I am certainly interested in doing elsewhere. Although it's worth noting that appealing to the complexity of this approach in this way will help defend the Kosslyn Model in Chapter 4.

So I will now introduce these components here as a suggested way that the basic offline-simulation flow diagram shown in Figure 2.1 above could be developed further. So on top of the four mechanisms introduced above, perhaps it would be useful to also introduce two further mechanisms that give this process meaning and also help to keep it running in an appropriate manner. These are respectively: an *interpretation capacity* and an *executive controller*. The details of the function of each of these six proposed functions is listed below, along with that of the main mechanism to be run offline:

1) **Simulation Motivator** – identifies when it would be beneficial to run a simulation and motivates its initiation: e.g. the decision to visualize a future situation, as it may be useful for planning; this may also involves the sending of instructions to manipulate the image to test out novel variations e.g. rotation, zooming etc.

[30] Although it is worth noting that Currie (1996: 249) did qualify this by saying that each component has a different level of plausibility from the other in that context

2) **Input Identifier** – chooses the appropriate inputs to be fed into the simulation process: e.g. beliefs and desires in the case of mindreading or decision making; or relevant memory traces in the case of visual imagery.
3) **Feeder Mechanism** – ensures the input is fed to the appropriate part of the main mechanism and that it is tagged appropriately as offline: e.g. signals sent with correct labelling to appropriate parts of 'visual buffer'.
4) **Mechanism to Run Offline** – a mechanism that has a normal online function, but which can be fed mock inputs and will give mock outputs that can be utilised for planning: e.g. the 'visual buffer' of the visual cortex which is used in both vision and visual imagery.
5) **Offline Output and Output Detector** – which allows the processes running the simulation to detect the outputs and keep them labelled as offline so as not to confuse them with real events. It also takes the output from the main mechanisms and processes it appropriately to generate the final outcome: e.g. visual images, pretend decisions or feedback of imagined bodily movement in motor imagery.
6) **Interpretation Capacity** – something that can use the outputs in a meaningful way that will benefit survival: e.g. interpreting a visual image so that it can be used to inform further manipulations or to draw some conclusions.
7) **Executive Controller** – controls the amount of time and resources used by the simulation system so that it doesn't affect real time performance: i.e. stopping imagery from occurring while urgent events occur in real time (e.g. stopping daydreaming while your house is on fire), and stopping iterations continuing beyond a useful point (e.g. prevents perseveration: which is the obsessive repetition of actions or thoughts).

Table 2.1 – Potential mechanisms Necessary for an offline simulation process

The above is aimed to cover the full range of mechanisms required for modern humans to deploy their imaginative capacity: from instigation to interpretation; and from original input stimulus to final actual response. The relevance and importance of these proposed mechanisms will be ascertained throughout the thesis. Interestingly this approach also offers a way that this process can be used in increments to support a step-like imagistically based reasoning process. This is perhaps best illustrated in the diagram below of roughly how I envisage these mechanisms being inter-connected:

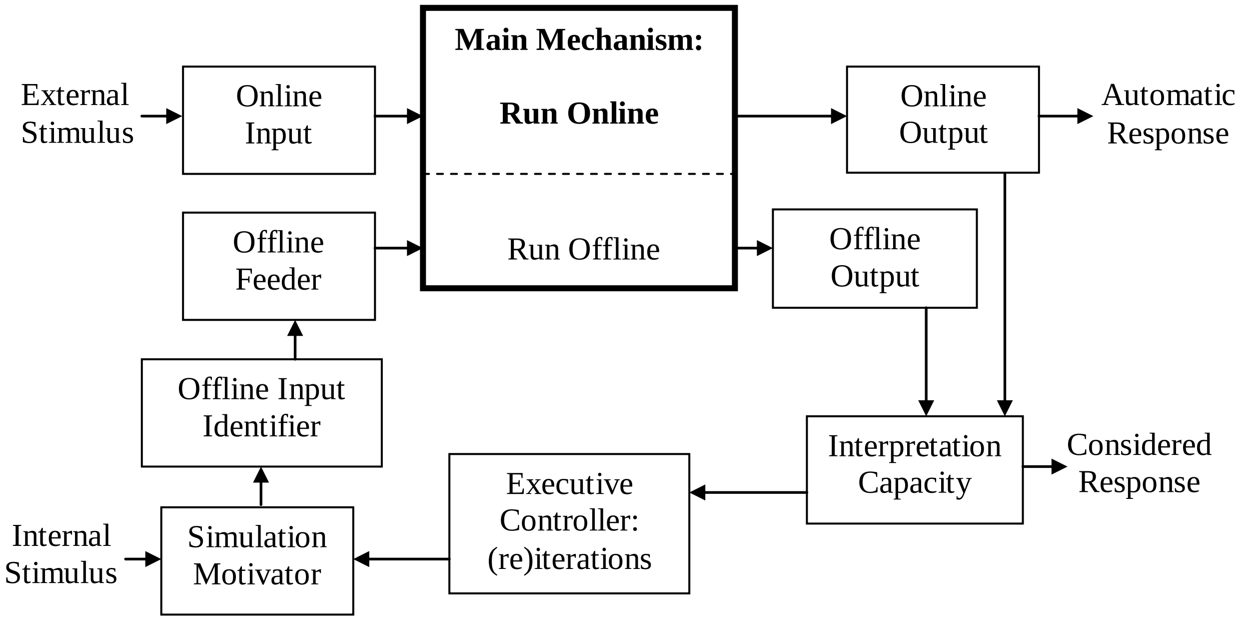


Figure 2.2 – Schematic of the proposed main components involved in full offline simulation (This is presented generically so as to possibly apply to imagination in any modality: but note that a visual imagery specific diagram will be given later)

Note that equivalents of the original components from the basic offline simulation represented in Figure 2.1 are still obviously included, but that these are labelled slightly differently (e.g. the offline input is now split into two mechanisms). Also note how the imagery episode can be run iteratively should the simulation motivator deem it necessary and the executive controller allow it. This might involve comparing the current state with a final goal state and ascertaining that further changes to the image are necessary to complete the task: for example, when ascertaining if two rotated objects are in fact the same shape. This process could also be terminated if urgent online issues become more important to deal with: i.e. when immediate actions are more important than planning and deliberation. Also note I have abbreviated the ‘Offline Output and Output Detector Mechanism’ from Table 2.1, to simply the ‘Offline Output’ in the Figure 2.2, since these will be treated in conjunction from now on (and see the intro to Part II for more on this).

In Chapter 4 I will try to identify potential candidates for each of these mechanisms in visual imagery based on the Kosslyn Model. Some will be more obvious than others and some remain poorly understood, or are a part of a wider analysis of how the mind works. Hence this diagram can be used as a platform to analyse the relation between vision and visual imagery. I would suggest that we have now discussed enough here to provide a solid introduction to an imagination-as-offline-simulation approach. We can now apply this by looking at how it relates to *emulation* processes.

2.3 – To Emulate or to Simulate? That is the question

In a recent article, Grush (2004) has put forward a theory that puts emulation at the centre of motor control and motor imagery. He even links it to filling in missing parts of perceptual experience and pre-empting what we expect to perceive visually: i.e. he appeals to a visual emulator. Hence he seems to be addressing two of the main areas of interest in this thesis: what are the underlying mechanisms of imagery and what is their relation to perceptual capacities. In his paper he spends some time explaining why emulation theory is to be favoured over simulation theory, or why simulation theorists may have meant emulators all along. He summarises his view on the differences between the two positions in the following quote:

> ‘To make an analogy: The emulation theory claims that motor imagery is like a pilot sitting in a flight simulator, and the pilot’s efferent commands (hand and foot movements, etc.) are translated into faux “sensory” information (instrument readings, mock visual display) by the flight simulator which is essentially an emulator of an aircraft. The simulation theory claims that just a pilot, moving her hands and feet around but driving neither a real aircraft nor a flight simulator, is sufficient for mock sensory information.’ Grush (2004: 384)

This is a serious claim and makes the simulationist responsible for quite a large error. But mixed up in this analogy are several assumptions that need to be unpacked, not least of all Grush’s claim that a flight simulator is ‘essentially an emulator of an aircraft’. So in order to reply to this criticism, in what follows, we will need to look at what he means by an *emulator* and what he thinks the simulationist is so mistaken about. I will eventually give a response at the end of this section that suggests that, while Grush makes a good point against simulationists in terms of what they fail to address specifically, it is a limited one, and we are probably not as mistaken as he suggests once we explicitly deal with his concerns. Hopefully then, once his use of the terms ‘emulator’ and ‘simulator’ are analysed and tidied up, much common ground can be found. The added benefit of the exercise is that it will further illustrate what the offline simulationist means and will identify a potential candidate for *full simulations* already being used in *offline simulation* processes in the human brain. So in order to begin this clarification process, some background on emulators is needed so as a basis for the following discussion on emulation versus simulation.

2.3.1 – Emulation basics and definitions

The original idea of emulators comes from factory processes that can be made more efficient by predicting probable future states of the systems involved and correcting for them before they happen. Hence the terminology involves a ‘controller’ that issues commands to a ‘plant’, which is the system to be controlled. The model of the system that can predict future states of the plant is called the ‘emulator’. This is illustrated the basic emulator schematic diagram below:

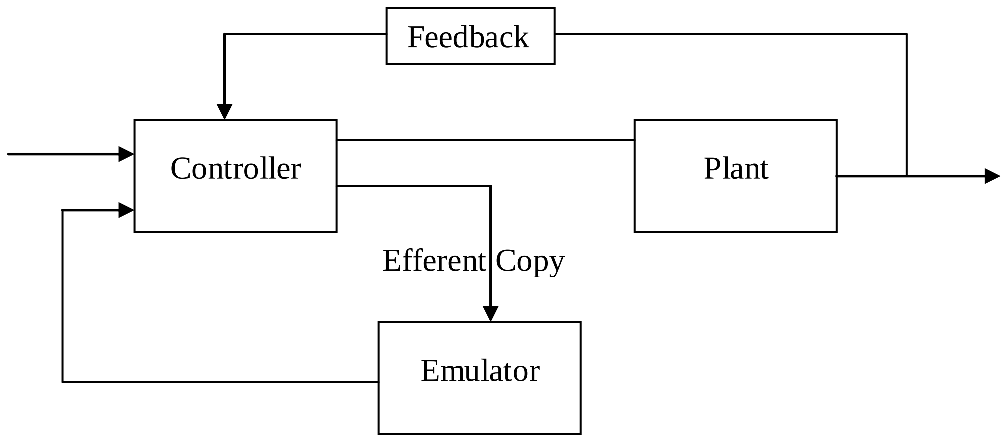


Figure 2.3 – Basic Emulator Schematic

In order for the emulator to track what is going on it receives a copy of the control signal called an 'efference copy' and also information about the current situation of the system (and other information to do with a complex 'Kalman Filter' that we need not address here). Based on this it can model the system and predict its future states. Usually this can be done at a much faster rate than waiting for feedback from the system itself. Hence the emulator can feed the controller quicker feedback and it can therefore adjust its signals with a finer grain of accuracy given the desired goal state.

Given the basic description of an emulation system above, we can maybe draw out two very important distinctions. Firstly, I would suggest that in order for something to qualify as an *emulator*, there must be some currently occurring real system that it is trying to emulate or mimic. This is because the whole purpose of the emulator is to provide fast feedback in order to correct the controller signal to make the plant work *more efficiently.* Hence its main (proper) function is to fine tune the activities of a currently occurring system within which it is embedded. This is what is special and unique about an emulator; it is part of a wider emulation system and there is a real world system that it emulates. I suggest then that the emulator gains its status in lieu of being part of an emulation system, and if it is not placed appropriately in connection with a plant and controller then it ceases to be an emulator strictly speaking.

With that in mind we can make the second point and ask what happens when we take the rest of the emulation system away (e.g. the controller and the plant)? And bear in mind that the controller has a dual function of: a) controlling the plant; and b) integrating feedback from the plant and the emulator in order to fine tune its commands to the plant. Here I would suggest that the function of the emulator in isolation is to: take a mock input; process it based on some rules about how the online version might work; and to supply a mock output that is a prediction of the output of the real system. But what we have just described is just a 'full simulation' process according to my categorisation above: i.e. the target system is completely simulated in a different medium, like weather predictions based on a binary computer simulation. So the definition of an emulator that I think will be most useful in what follows, is this:

> **Definition - Emulator**: An emulator is a *full simulation* process that is embedded in an emulation system as its primary function. An emulation system normally utilises an emulator to provide quick forward-modelling feedback in order to help it fine tune the actual control of a *currently running real world system*. Thus the primary function of an emulator is to work within an emulation system and to mimic a currently active target system.

I make these distinctions because in what follows we will discuss using an emulator offline to underpin 'motor imagery'. Now as long as the primary function of the emulator-as-full-simulator is to work in an emulation system then it is correct and proper to refer to it simply as an 'emulator'. However if it is removed and permanently isolated from that primary function, then I suggest it should return to being referred to as a 'full simulation' mechanism (or simply a simulator; which as it happens used to be an emulator). Hence the emulation title is inherited from the system that it is currently being used for, or indeed has evolved in. But its isolated function within that system is usually to provide a 'full simulation' of some real world system (the plant) in order to aid the controller.

To illustrate how this distinction immediately helps, lets refer again to Grush's claim that a flight simulator is 'essentially an emulator of an aircraft'. For this to be true given the above definition, the flight simulator would need to be providing feedback into a real aircraft system in order for the controller to fine tune (not even control) operations. Without this it is just a simulation of flying a real aircraft: which is why it's called a 'flight simulator' and not a 'flight emulator'. So I would suggest that Grush has extended the use of the word 'emulator' beyond what I think it is best reserved for, as per the definition above[31]. This may just boil down to semantics, and in that case nothing much rests on it, apart from possible confusion. The simulationist can't stop Grush using the term 'emulator' for what we think of as 'simulators', and we can keep this in mind when reading his work[32]. The important thing is that we are getting clear about what the function of these mechanisms are and how they may relate to each other and our wider activities. I think this initial clarification may explain some of what Grush is complaining about. However we still need to explain where he thinks the simulationist has gone wrong. This will involve explaining where the emulator fits into motor control and motor imagery.

2.3.2 – Emulation and motor control

In terms of motor control, Clark and Grush (1999) give an account of how limb movement could be emulated in order to fine tune and smooth our reaching and grabbing actions. The

---

[31] A syndrome I think Kendall Walton once referred to as: 'have theory will travel'. Something a simulation theorist must also be in self-regulatory guard against i.e. applying simulation to everything that on the surface resembles it but may be inappropriately labelled on further reflection (quoted in Currie 2003).

[32] And see Hurley (2008: 21 – footnote 3) for a similar point: 'I, like many others, use "simulation" to include "emulation" in Grush's sense'. This allows her to use the word 'emulation' for occasions of social learning, where for example, a child emulates or copies what an adult is doing.

need for this emulated input is to provide faster, locally generated (intra-brain), projected feedback, rather than having to wait for distally generated slower feedback from the limbs or the environment to arrive:

> '[The emulator] allows the system to exploit mock feedback signals available ahead of the real-world feedback, and hence allows rapid error-correction and control. It can support reasonable sensible behaviour in the total absence of real-world feedback.' Clark and Grush (1999: 6)

They explain that this extra process, rather than being an expensive hindrance, has a beneficial function. That is it has the overall effect of more rapidly correcting and smoothing the trajectory of the desired movement, hence resulting in improvements in the execution of a real action. So the utility of emulating motor control is fairly evident, as is its evolutionary benefit. Perhaps we can now make the analogy that: the motor control areas in the brain are equivalent to the controller; the movement of limbs is equivalent to the plant; and bodily feelings are the feedback from the system. This is summarised diagrammatically below:

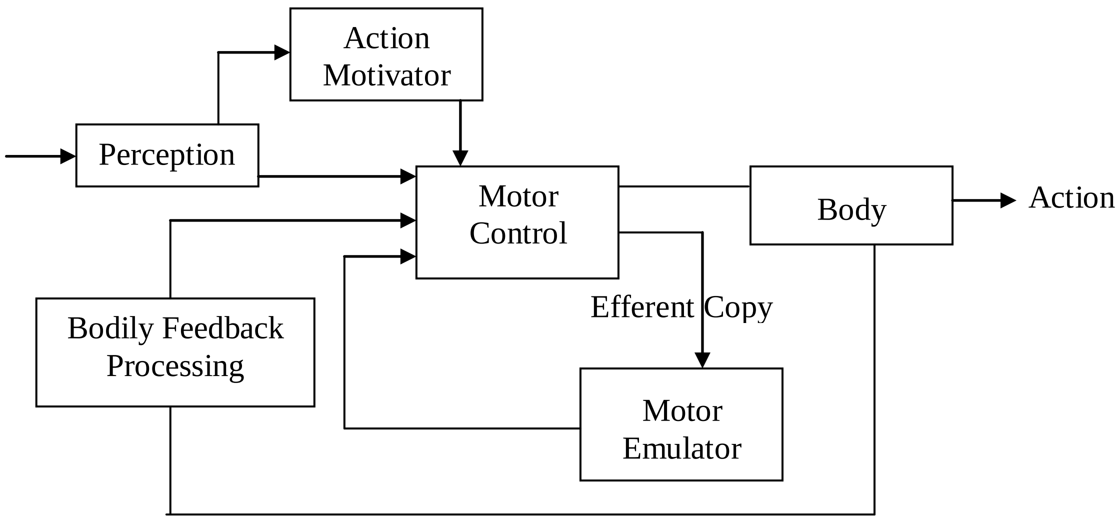


Figure 2.4 – Emulator functioning in an online motor control system

I have placed perception on this diagram as feeding the 'Action Motivator' box and going directly to the 'Motor Control' box, because it is thought that visual perception is used directly to help guide motor control, as well as to be used more obviously in motivating our actions. The use of perception in action control is the basis of the idea that action and perception share information and affect each other dynamically: i.e. in sensorimotor control circuits (Hurley 1998, 2001 – and see later). This is evident in reaching and grasping movements where you usually watch what you are doing to help guide that action (just try it with your eyes closed to see how much it helps).

Given this motor control context, we might perhaps now want to ask: what actually is the motor emulator itself? Grush (2004: 379) offers two alternatives and favours the second of his own formulation. So although he admits there are other possibilities, we only really need to discuss the one he favours in detail order to analyse his critique of simulation theory. The first alternative he mentions is an 'associative memory look-up-table' where: 'previously observed musculoskeletal input-output sequences' are stored; and on receiving the input signal the closest associated output is produced. Call this a '**Memory-table Emulator**' and we can see this also as a crude simulator that yields the most likely output given any set of inputs based on previous experiences. This feature seems to anticipate some work we will analyse in the next chapter in terms of visual imagery using memory traces to determine inputs to the visual buffer. But for the reasons just stated, and following Grush, I put this option aside for now.

Grush's favoured emulator theory is one he developed in his Doctoral dissertation (1995), which he calls an '**Articulated Emulator**'. I will give some brief details of this in the following, but this is only aimed to be just enough to show the reader this can also be interpreted as a full simulation process. The real system the Articulated Emulator model is the musculoskeletal system and its various articulated limbs and joints. The feedback from the musculoskeletal system is kinaesthetic and proprioceptive body reactions (hereafter termed - *bodily feedback*), which are measured by receptors in the limbs. This system is modelled in the brain by the Articulate Emulator by representing equivalent components (or articulants) for those of each real limb and joint, which are sorted into neuronal groups 'whose firing frequency corresponds' to these different parts of the body and where they are located (e.g. bent, extended, rotated). These representational neuronal groups interact with each other in an excitatory or inhibitory fashion to model, or predict, the real musculoskeletal positions. Then (and this bit is quite vague) just as the real musculoskeletal movements are detected by receptor to provide bodily feedback: 'the articulated emulator can have a "measurement" taken of the same variables and thus yield a mock sensory signal.' [pg.380]. We will return to analyse this 'mock sensory return signal' later, because the apparent omission of this mock output is the main gripe Grush has with simulation theorists.

May I suggest that, in this context, the details and the feasibility of the internal workings of the Articulated Emulator are not as important to focus on, as how it works in the imagery case; I assume anyway that this is an in-house problem for emulation theorists to sort out. So I will not analyse the process by which it works, but only at this point suggest that both types of motor emulator described above, seem to be a separate and independent brain mechanism, that somehow model, or mimic, the real system. And given what we have discussed above, it therefore seems that Grush's Articulated Emulator is a full simulation of the working body based on neuronal computations. However it is still properly called an emulator because its specific function is to help optimise a currently running real system; i.e. to smooth body movements. In slogan form perhaps we can say that:

In normal motor control: the emulator is a full simulator embedded in an emulation system.

These points will become relevant below.

2.3.3 – Emulator and motor imagery

Grush then suggests that in motor imagery the motor emulator operates as follows: '…real sensory information has no effect. The emulator's state is allowed to evolve according to its own dynamic and efference copies, if any; there is no "correction" or alteration from the senses …[and] the motor command must be suppressed from operating the body.' [pg.384]. I interpret this as suggesting that the emulator is cut off from real motor feedback and left to run according to its own dynamic, and the controller prevents the emulator from affecting its output commands to the body. I take this to be roughly the equivalent of taking it offline from its normal online activity of smoothing movements.

So when the emulator is used in imagery it would seem that the emulator itself is taken offline and given mock inputs in order to simulate possible body movements that are currently not occurring. These mock inputs are no longer efferent copies because they aren't a copy of real signals that are being sent to the body. And the imagery system is not an emulation system as there are no real movements to mimic or smooth. On receiving the mock inputs, the motor emulator processes them and converts them to signals that the body may have returned and this is what our imagined feeling of body movement is *based* on according to Grush. Given this, to me it seems clear that the emulator is detached from its normal role in an emulation system and is being used as an offline simulation of body movement. This seems to be an unusual combination of taking an emulator, which in itself is a full simulator embedded in an emulation system, and using it in an offline simulation process to yield motor imagery. And while this may sound odd it does seem to reflect the situation, and the distinctions made earlier seem to be able to cope with this proposed dynamic. For example: it is still technically an emulator as this is its primary evolved function in the real system. However, in the imagery system a part of the online emulation system (i.e. the emulator) is fed mock inputs and returns mock outputs that yield information equivalent to what may have been felt had the inputs been real; which is just the definition of an offline simulation given above. I think this way of talking fits with how Grush talks about taking a *visual-emulator* offline:

> ‘An emulator by itself does not decide when it gets operated on-line versus off-line. Presumably there is some executive process that makes use of emulators, sometimes for imagery, sometimes for perceptual purposes.’ Grush (2004: 389)

So the use of the offline and online distinction is explicit in his writing. And we can perhaps suggest that when undertaking an episode of motor imagery, it does not get used to fine-tune any currently running system, but rather to help plan and eventually execute future actions more accurately and efficiently (and see later for how doing this covertly might give us an evolutionary benefit). Hence, perhaps it is more appropriately referred to as an offline simulation process while it is being used in this function. Maybe then we can also summarise motor imagery in slogan form:

> **In motor imagery:** the emulator is a full simulator utilised in an offline simulation process.

I think this approach makes for a clear and useful explanation of what may be going on, and can be helpful in resolving Grush’s complaints about simulationists.

Before doing that, it may be useful to can observe that in the quote above Grush mentions executive processes and this may neatly link with my Executive Control mechanism mentioned in Section 2.2.3. The best guess at the moment is that these ‘executive processes’ are generated somewhere in the prefrontal cortex (PFC), and this has some general empirical support[33]. Also consider this quote for an indication that this interpretation is on the right lines:

> ‘Dorsolateral prefrontal cortex is also notoriously activated during preparation states, when a decision must be taken about which finger to move or about when to start a movement (Frith et al. 1991).’ Jeannerod (2001: S106)

This also matches up with work by Kriegel’s (2007b) on the subject of the neural correlates of consciousness. There he elaborates on how the prefrontal areas might communicate with more (downstream) posterior areas, in a process that yields these kinds of phenomenally conscious representations of: body movement, visual perception and their imagistic counterparts. I will return to this topic briefly in Chapter 7.3, but for now this gives us an initial useful example of the function of the Executive Control mechanism depicted in Figure 2.2 that can be referred to later.

So returning to analysing the role of the motor emulator in motor imagery, we can now return to look at how the rest of the motor imagery system functions apart from the emulator. Here Grush (2004: 383-385) agrees with the simulation theorist, that real online motor control areas in the motor cortex are taken offline in order to generate mock efferent (outgoing) motor commands; he calls these areas the efferent control centres. Perhaps we can also suggest that this is preceded by a conscious decision to undertake an episode of motor imagery, and most probably this is in order to solve a physical problem or practice a physical task. So after this, these mock commands to move are sent to the motor areas that also control the real commands (i.e. the controller). These outgoing commands (inputs) would normally travel to the limbs (the plant) and movements would be executed, but in

[33] C.f. Faw (2003) and Russel (1997).

this case overt movements are (mostly[34]) suppressed in imagery. So instead, according to Grush, the mock commands are only sent to the emulator, which processes these as normal, and then returns outputs that retain their 'mock' imagery label.

Furthermore Grush agrees that brain areas that normally receive bodily feedback (e.g. the cerebellum) are also active as the simulationist suggests. But he disagrees that just taking the areas that normally detect the return signals offline, is enough to explain how the offline *input* and *output* areas are linked together. The **important** thing to realise is that in the real case, actual movements are themselves normally detected by: '…stretch receptors and Golgi tendon organs' in or near the muscles [pg.385]. These signals are what is returned to the cerebellum (at the base of the skull) and are what is thought to underpin your feelings of bodily feedback (*reminder*: bodily feedback is short for proprioception and kinesthesis). Grush goes to some length to emphasise the difference between *outgoing motor commands* causing movements and *returning bodily feedback* that have detected these movements in the muscles.

Grush's main claim then is that this is where the simulation theorists go wrong. They just assume taking the relevant motor areas offline will provide the same faint phenomenal sensation of body movement. However the actions need to be realised in muscles and detected in movement receivers for this to happen in the real case, and simulationists don't offer a parallel offline way of getting this return signal. So it is this second returning part of the process that Grush feels is neglected by the simulationist who usually only talks about taking the efferent motor command areas offline (e.g. Jeannerod 2001). Which would be the equivalent to sending out loads of mock commands and receiving nothing back. Hence his analogy of the pilot flapping their hands about on the controls (mock inputs), but these

[34] There is some inefficiency in this process as some slight movement can occur as well as other physiological changes - see Jeannerod (2001: S105)

not giving any useful return mock outputs on the dials of the cockpit. Grush thinks therefore what is needed is an emulator to convert the mock inputs to mock outputs.

In reply to this criticism of apparent omission of a crucial step, the simulationist can simply admit that if it is true that taking the motor command centres offline only simulates half the process, then this is indeed an error. This is because the simulationist would certainly not want to propose a simulation system without any mock outputs; indeed this would prevent it from being a simulation almost by definition. However I offer the following quote from the simulationist literature that shows that perhaps we may not be as confused as Grush thinks:

> 'If the emulator hypothesis is true, then motor instructions that are prevented from being transmitted to the limbs might still activate the emulator, the operation of which then gives rise to those processes we recognize as the perception of bodily movement.' Currie and Ravenscroft (2002: 84)

From this quote we can clearly see the emulator is being envisaged as the translator from 'outgoing mock movement commands' to 'returning detections of mock movement', pretty much in the same way that Grush envisages it need to. So Grush may have simply been unaware of this fairly sympathetic inclusion of the emulator in simulation-based explanations of motor imagery.

We might also want to say that his criticism of the simulationist is a little bit unfair, since both Jeannerod (2001) and Johnson (2000) (his main targets) may refer to areas of the brain that are responsible for action detection, which are also active during imagery. For instance, Jeannerod (2001: S105) talks about the cerebellum being activated during imagery, and Grush (2004: 385) cites this as the most likely neural correlate of a motor emulator. So perhaps they think that just by listing evidence for distributed overlapping activity over *most of* the areas involved in the *motor system*, then they are also including those responsible for ‘bodily feelings’. All this would then imply is that they have failed to give an explicit account of how the mock action commands are translated to mock bodily feelings. So it may have just been an omission on the simulation theorists part to be *explicit* about something as simple as, for example, a Memory-table Emulator, which would be needed to translate the ‘mock commands’ (mock inputs) into ‘mock bodily feedback’ (mock outputs). If both these areas are active during motor imagery, then it seems there may just be an *implicit* assumption that one is *somehow* being related to the other in order to simulate the complete experience.

What Grush *has* done is offer the simulationist some emulation-based details of how this translation might be made without having to posit a completely new and specifically designed mechanism. We can simply process mock inputs and outputs via an emulator mechanism that is already doing this for another purpose. Hence offers us a way of making this part of the process explicit and giving it a parsimonious solution. I think this makes some sense of why the simulationist may have assumed this process, without realising it needed a more detailed explanation.

So in conclusion to this section, perhaps I can now explain how the simulationist can respond to Grush’s complaints that I introduced at the start of the section by way of the flight simulator quote. Given the above, perhaps we can accept that in an analogous way to motor imagery, the simulationist might have failed to explicitly mention how the input from the pilots interface with the controls, are translated to meaningful outputs as readings on the cockpit display. So we can *agree* with Grush that the simulationist still owes an explanation of how these mock inputs are converted to mock outputs. But the completion of this explanation can be as simple as using a *full simulator*, which could be running on a computer that is hooked up to the cockpit and programmed appropriately to do this job. That is, the computer can: firstly receive the signals coming from the cockpit controls; then process them using some programmed rules about how they relate to each other normally in the online situation; and finally these processed signals can then be sent to adjust the cockpit displays appropriately. This inclusion of a *full simulator* parallels the equivalent appeal to a motor emulator that the simulationist might use to explain motor imagery as described immediately above.

Conversely the simulationist can still *disagree* with Grush that this constitutes any need to re-name this process as an emulation of an aircraft, since in this case the *full simulator* is not being used for any other online fine-tuning emulation process. Hence it is not strictly speaking best referred to as an emulator and it is still appropriate to call it a flight simulator as per normal language usage. More specifically the mechanism we have now introduced isn't *necessarily* being used to emulate and fine-tune an online real mechanism (i.e. a real plane flight) and potentially all that is needed is a *full simulator* to translate the inputs from the cockpit controls to related outputs on the cockpit displays. Hence there is no need to appeal to an 'emulator' in this process at all and we can just talk in terms of simulations of different varieties. The simulationist can now counter-accuse Grush of using his emulator terminology too broadly and simply trying to insert the term 'emulation' where the term 'simulation' is already doing an adequate job.

I think the confusion arises because there is a crucial disanalogy between the flight simulator and motor imagery explanations. And this arises at the point where Grush is suggesting a motor emulator is available to take offline to help generate motor imagery. But no such mechanism is suggested (yet) in the flight simulation case and hence there is no need to use emulation terminology at all in this situation. However, if Grush now wants to suggest that there is an aircraft emulator on board, whose primary functions is to fine-tunes the online control of a real aircraft flight, then the simulationist can also parsimoniously appeal to this mechanism in the offline case, and perhaps now it is more appropriate to call it a flight emulator that is being used offline in a flight simulation process.

Hopefully the distinctions and clarifications I have given above have made it clear when I think it is most appropriate to use the term simulate or emulate; after all that was the original question in this section title. But putting definitional issues aside for now, perhaps the main point is that we can accept that Grush's demand for clarification on this point in terms of mental imagery, yields a useful elaboration of to the simulationist approach. It also offers another useful mechanism for simulation theorists to adopt in their future explanations. It seems to me that both theories can support each other, since in some sense we are both in the business of suggesting simulation-style processes are useful in explaining certain aspects of our mental imagery and action control.

So in summary, this section has aimed to do three main things. Firstly it has responded to Grush's criticism of simulation theory as described immediately above. Secondly, it has also elaborated further on how an offline simulation process might invoked to explain motor imagery and how it might be able to utilise an already existant motor emulation mechanism. This analysis has also involved giving certain clarifications about the similarities and differences between emulation and simulation and I have developed some useful definitions that I will utilise in what follows. Thirdly, this may now allow us to look for parallel explanations in the field of visual imagery. In this context we can now ask whether an already existant 'visual emulator' can be incorporated in a simulationist explanation of visual imagery; this will be discussed further in Chapter 4. For now, I would suggest that the exposition above allows the Simulationist and Emulationist approaches to compliment each other once again, and for the definitions given to avoid confusion in future. In order to test this out further, perhaps we can now look at how Hurley (2008) invokes a combination of emulation and simulation in order to underpin a mental function she calls *deliberation*.

2.4 – Simulation, Emulation and the Shared Circuits Model

Before continuing, please note that Hurley (2008) refers to Grush's 'emulator' as a simulator so as not to confuse it with her use of the term 'emulate'. She uses emulate to refer to *action copying* of a certain kind in the social cognition arena. I will therefore use a capitalised 'E' to refer to Grush's *Emulator* mechanism from now on, or just refer to it as a simulator when it is not used in guiding actual movements as described above. Also note, I am not attempting to give a full exposition of Hurley's position here. I am just introducing the crucial points in order to illustrate a parallel approach that relies on many of the same processes. This is aimed to provide support for my approach generally, as well as elaborate further on the potential nature and function of STIm.

That said, Hurley's (2008) Shared Circuits Model (SCM) aims to show how complex behaviour, like mind reading, planning and social cognition, can be built up and enabled through five hierarchical layers of basic processes. Her main focus is seeing how much of this work can be done by the interaction of simulation processes, control processes and mirror neurons[35]:

[35] Mirror neurons are areas of the brain that fire both when we do an action and when someone else does it. In a similar vein *canonical neurons* not only fire on seeing an object that affords a certain type of action, but also fire when the subject performs that same afforded action e.g. seeing scissors and cutting with scissors. But note that care must be taken when claiming too much sophistication on behalf of these mirroring neurons alone, on the understanding that their activity needs to be imbedded in a system that functions to interpret this very basic synchronised firing activity (cf. Rizzolatti 2005).

> ‘ The [SCM] shows how subpersonal resources for control, mirroring, and **simulation** can enable the distinctively human sociocognitive skills of imitation, **deliberation**, and mindreading. The model has intertwined empirical and philosophical aims.’ Hurley (2008: 11 – emphasis mine)

Each layer of processing she invokes builds on the one below, with Layer 1 being the simplest and Layer 5 potentially sub-serving our full mindreading capabilities. As it turns out she pretty much uses Grush’s (2004) Emulator-as-simulator, as a *basis* for her arguments for how we plan, predict and understand actions in others and ourselves. So once again we perhaps have a theory that makes use of an Emulator as well as an ability to *re-u*se this mechanism by taking it offline[36]. We will therefore see below that the two layers we are most interested in this context are layers 2 & 4: where layer 2 is simply Grush’s online Emulator involved in online motor control; and layer 4 allows *deliberation* by taking the Emulator offline and allowing motor imagery. I wont say much more about Layer 2 as this has been discussed above. Hurley also introduces the concept of reverse simulation, which gives you the relevant inputs that any given set of outputs might have been caused by; which is the opposite way round to normal simulation[37].

Before analysing certain implications of the SCM model, we need to make explicit an assumption in Hurley’s work, which was mentioned briefly earlier. That is, there is a certain amount of overlap between action perception and action control. Simply put, we use vision to watch our own actions (e.g. hand movements) and use this, along with an Emulator, to help our motor system guide and smooth those movements. Hurley refers to this as ‘sensorimotor feedback’, which includes motor feedback (bodily feelings) and

[36] Recall I quote Hurley (2008) earlier as defending as re-use style of simulation that I think is roughly equivalent to an offline simulation process.

[37] This could be explained by something as simple as just reversing the Memory-table Emulator process, but again I won’t have time to deal with this interesting mechanism invoked in SCM in detail. Investigating the plausibility of this mechanism in humans is another potential avenue for future analysis. And see Miall 2003 who suggests some other uses for reverse simulators in artificial systems.

sensory feedback from the environment e.g. vision and sound. In addition these areas may interact as a unified whole (what fires together, wires together) making a strong link between perception and action as studied in control theory; hence it is referred to as a shared circuits model. It is possible that if this conjoined system is then used offline we may see some parallels between our visual and motor imagery; which is generally what's found (c.f. isochrony studies reviewed in Jeannerod and Frak 1999 & Kosslyn et al 2006). Analysis of the visual part of this, and its contribution to visual imagery, will form a vital part of the next two chapters.

We can now look at the implications of Layer 4 to this thesis in a bit more detail. To put this in context, immediately below is a tabulated summary of the five layers that Hurley identifies and their relevant functions (c.f. pgs. 19-20):

Layer	Function	Mechanism	Comments
1	Online Motor Control	Motor Control of Body	Feedback is from environment only
2	Offline Motor Prediction and Control	Simulation (Grush's - Emulator)	Improves action execution. Efference copy also potentially enables distinctions between self and world activity
3	Other's action and goal detection and copying	Mirroring and reverse simulation	Takes detected mirrored 'motor outputs' as inputs to reverse simulation in order to prime, emulate or imitate
4	Inhibition of overt action for discrete planning	Monitored output inhibition combined with simulative prediction and simulative mirroring	Allows discrete (not overt) consideration of possible actions and understanding other's actions and goals
5	Counterfactual consideration	Counterfactual input into reverse simulation	Allows other's counterfactual actions to be analysed, as well as reaction to your possible actions

Table 2.2 – Summary of Hurley (2008) sub-personal functional layers in the SCM

2.4.1 - Layer 4 – "monitored output inhibition combined with simulative prediction and/or simulative mirroring"

In terms of the function of Layer 4, I will focus mainly on its role in supporting *deliberation* since it is the most related to STIm. Deliberation is the ability to simulate alternative possibilities whilst preventing yourself from acting them out overtly. This latter feature is what Hurley refers to as action inhibition, and deterioration in this ability can have debilitating action control consequences (see below). Therefore this stage importantly involves the inhibition of these imagined and/or imitated actions from being actually executed. At this stage it becomes possible to test out future potential actions that might be appropriate given the current affordances in the environment and to copy other's actions in a motor imagery episode. This is mostly just equivalent to utilising motor imagery to rehearse movements, where the Emulator is run offline as described in Section 2.3. More interesting though is what Hurley thinks this is used for and this is well represented in this quote below. Note she also uses very similar language to that quoted from Dennett (1995) in Section 2.2.3 above:

> 'Multiple simulative predictions could provide information about results of alternative possible actions, rather than anticipating results for ongoing actions. Simulated results of alternative possible actions could be compared for the closest match to a target prior to actual action. Layers 2 plus 4 could thus provide information for "trials and errors in the head" (Millikan 2006) prior to actual trials and possibly fatal errors, allowing simulations to die in a choosers stead. They could thus enable counterfactual instrumental deliberation and *choice among alternative possible actions*.' Hurley (2008: 16 – my italics)

I think this provides confirmation that the remit of deliberation is essentially the same as 'strategy testing' as per my usage. So I would suggest a strong link to my offline simulationist approach described above is evident.

However Hurley goes one step further here and joins this together with her action–reading, imitation and simulation capacities. These other capacities are mainly based on mirror neurons and reverse simulation processes that I won't go into here in detail[38]. Suffice to say these features are developed at the same middle functional level of description as that of Emulator and offline simulation systems. Hence indicating the potential of combining an offline simulation approach with mirror neuron theory in order to explain some other useful human abilities. Hurley suggests that this combination of mechanisms allows for covert imitation and learning from other's actions. Initially it seems obvious why sometimes inhibiting our natural tendency to copy other's actions has an evolutionary benefit, since we don't always want to reproduce what we see happening: e.g. people having accidents. But this process also yields another advantage to 'internal imagery', which I had not anticipated before reading this paper. That is it prevents your competitors copying useful techniques from you, or knowing what you are copying from them:

> 'Imitation with selective inhibition has the advantage of theft over honest toil: Instead of letting *hypotheses* die in his stead, a selective imitator lets *others* die in his stead, reaping the benefits of success without unusual native wit while avoiding the cost of trial and error. Imitative social environments may in turn generate pressure to prevent successful techniques being appropriated cost free by competitors, resulting in capacities for covert or simulated action, shielded from potential imitative theft.' Hurley (2008: 8)

In other words covertly rehearsing actions for future plans using mental imagery, and imaginatively imitating and learning secretly from others' actions, has another obvious evolutionary benefit. This could then provide an additional drive to develop internal ways of testing strategies without actually risking trying them out, or revealing what you know about them to others.

---

[38] See footnotes 35&37. But very briefly, the main idea is that mirror neurons imitate and generate the output commands that other people's observed actions stimulate. These can then be put through a reverse simulation process that yields what the relevant inputs might have been. These can then be imitated by actually implementing them, or merely simulated by imagining doing them. This allows us to learn from them or to help understand what these actions might have been intended to achieve. It is a short step from here to understanding the intentions of others in terms of a basis for mindreading: in short, you can work out what you might have intended had you been doing the actions you observe other people doing.

So given the above, perhaps we have once again identified a theory that, like mine, leans on simulation processes quite heavily. It also anticipates some novel benefits of simulation based mental imagery. This is because it allows clandestine copying and rehearsal of useful actions, as well as planning for future actions based on choosing between possible alternatives using mental imagery. Hurley is also very specific about how STIm-like deliberation abilities could have evolved in a hierarchical manner and how they may function more widely in our mental economy. Hence providing a parallel plausible story of how they may have evolved and how they may currently operate. I would suggest this provides some further support and motivation for pursuing this style of approach in what follows; which was mainly all this brief analysis of SCM was aimed at doing.

2.4.2 – Relevant extracts from other layers in SCM and peer commentary review

It may be worth briefly pointing out two other relevant implications of Hurley's SCM before a giving a brief review of the reaction it has had in the literature so far.

i) Layer 2

In terms of Layer 2, Hurley also discusses the utility of having an efference copy of the motor command just issued. Along with Frith (1992) and Currie and Ravenscroft (2002), she suggests that this is a good way of distinguishing actions that you caused (motor commands), with their expected emulated results, and those caused by the environment[39]. Put simply, if your actual movement is very different from expected, then it likely something unforeseen in the environment has moved it: e.g. I can feel my hand moving, but I didn't command it to, therefore there must be something moving my hand. Loosing track of this ability to discriminate self generated commands, with those caused by the environment, could lead to mistaking certain forms of imagery for perceptions: e.g. in

[39] Hurley (2008: 13) illustrates this with the train station illusion, where it is difficult to decide whether it's your train, or the adjacent one that is moving, without any other sensory cues

schizophrenic hallucination you may think you are hearing voices, when it is actually the 'voice in your head' which you are loosing a grip on.

ii) Layer 5

Layer 5 involves the counterfactual input of other people's possible actions in order to plan your own possible re-actions and consider the results: in other words you can deploy Machiavellian Intelligence[40]. According to Hurley [pg. 19] this underpins our ability for full-blown mindreading and marks a transition from: 'playing against nature' vs. 'playing against each other'. This perhaps indicates how offline simulation can be seen as necessary (but not sufficient) for both these processes. Again other parts concerning layer 5 mostly deal with mindreading and, though fascinating, lies outside the focus of this project. However this layered approach does support my suggestion earlier that mindreading involves a more sophisticated level of strategy testing and STIm. And it has been suggested elsewhere that predicting the actions of others, and keeping track of these complex volitional objects and their social interactions, can sometimes be more pressing and difficult than simply dealing the inanimate objects in the environment. This is the basis of the idea that social intelligence may have been a major evolutionary force in developing bigger brains (relative to body size), and hence is a major driving force in developing more powerful minds (c.f. Dunbar 2000, Humphrey 1976). This offers another way to link STIm, with what I called Mentalizing Imagination in Chapter 1 (i.e. imagining from another people's perspective or mindreading). This perhaps also indicates how sub-personal offline simulation processes may perhaps underpin both these mental capacities.

---

[40] C.f. Byrne and Whiten (1988, 1997).

iii) Peer commentary review

In terms of the peer commentary on Hurley's paper, a few themes were recognised and replies given by Clark and Kiverstein[41]. Many themes were to do with 'Imitation and Mirroring' and dealt with aspects of mindreading, so were less relevant here. In places (R3: pgs. 44-45), there was some explanation of how SCM could potentially explain certain aspects of autism spectrum disorder, by predicting it might be due to damage to the interactions of layers 3,4,and 5. I think this is close in *spirit* to imagination-as-offline-simulation based explanations of autism developed in Currie and Ravenscroft (2002: Chpt.7) and Currie (1996). One of the benefits of this kind of approach is that it may be able to explain both the executive function deficits and the mindreading deficits involved in autism, in one unified 'disorder of the imagination': i.e. as an inability simulate different possible situation and thoughts in order to plan and solve problems. Added to that, and linking it to Hurley's work, it may now also suggest a related way to explain perseveration and repetition as uninhibited affordance and imitation control[42]. But we must be wary here, as there are still other autistic symptoms that can't be so easily explained by this approach e.g. sensory hypersensitivity[43]. So there seems to be some further initial indications of potential for lining up SCM with other aspects of the imagination-as-offline-simulation debate.

---

[41] This is because sadly Susan Hurley passed away before this article was published. Both authors were collaborators of hers on the SCM and hence in the best position to predict what her replies may have been. I attended a conference in memory of her in Bristol where many of these themes were touched on (March 2009).

[42] Hurley (2008: 16) links this with imitation and utilization syndrome – where the subject can't stop imitating other people or acting on the affordances of an object in the environment. Perhaps the latter could superficially look like perseveration.

[43] And also note the warning by Boucher (1996: 229) that autism may be a complicated spectrum of different related disorders rather than one that can be solved with a single neat root cause. I have also written on this subject previously in my BA & MA (King 2002, 2005b), where I argued that autism-as-a-deficit-of-imagination explanations are fairly elegant and have good explanatory power, up to a point. It might be interesting to apply some of the methods that Kosslyn uses to defend his theory, by identifying differentially affected and spared mechanisms, to explain a range of related autism syndromes (see Chapter 4).

In terms of the interaction of levels 2 and 4, it should be noted that this underpinning of the ability to consider possible future actions drew very little criticism in the peer commentary. This either means there was less interest in this area, or this was seen as less controversial. If the latter is true, then it seems that SCM may have been fairly well accepted as a good basis for strategy testing within this context. If it’s the former then it seems its fate is linked most closely to arguments for and against Grush’s Emulation theory. So given this regrettably brief analysis of SCM, lets now apply what we have learnt in this chapter about emulation vs. offline simulation to the nature of visual imagery in the next two chapters.

3 – Defending Depictivism within the Visual Imagery Debate

'The biological problem in designing the psychology of organisms is thus to assure, as much as may be, that modes of representation are optimally mapped with kinds of tasks. People are one sort of solution to this problem… The key appears to lie in flexibility. Human beings apparently have access to a variety of modes of representation and can exert a rational control over the kinds of representation they employ. That is: How the available representational resources are exploited in any given case depends on what the agent takes the exigencies of the task at hand to be.'
Fodor (1975: 194) – showing sympathies for invoking different representational styles used in parallel in the brain

In the last chapter we looked at the kind of features we would expect an offline simulation process to exhibit *in general*. We also looked at examples of this from other areas of the simulation debate, such as in decision–making and motor imagery. This gave us an initial idea of the crucial aspects we should be looking for in this kind of approach. Given that analysis, perhaps we can suggest now that the most important thing to do initially in terms of visual imagery, is to identify a *main mechanism* that may then be available to be taken *offline*. To that end, in this chapter we will start to look specifically at whether the visual buffer is a good candidate for this main mechanism, and how the guidelines discussed in the last chapter might apply to vision and visual imagery. A crucial part of this process is to get clear on the way the visual buffer is thought to work within a theory put forward by Kosslyn et al (2006), which I refer to as the Kosslyn Model (as first introduced in Chapter 1). This is because it will be our main candidate for an online main mechanism that can be fed mock inputs to yield mock outputs, and hence support visual imagery.

Although I have briefly introduced some features of the visual buffer in Section 1.3, we will return to look at it in much more detail here. I will also try to defend the idea that *hybrid depictive representations* are the best way to characterise the way that visual images are processed within the visual buffer. This kind of representational style is characterised by being depictively organised, but also containing information in a more propositional format, and hence it is referred to as a hybrid depictive representation. What this means will be explained in detail in this chapter and this is important because it is crucial to understanding how the visual buffer can potentially be used both for online vision and offline visual imagery. Much of the analysis below will take the form of explaining and scrutinizing what has become known as the ‘Imagery Debate’ (c.f. Tye 1991). This dialectic puts *Depictivists* up against *Propositionalists* and this was also briefly introduced in Section 1.3, but will be expanded on in more detail in what follows. I will defend the position proposed by the Kosslyn Model that suggests that it is a specific combination of these two approaches that best describes what might be going on in terms of representational styles in the brain. The aim of this chapter is to show that a hybrid depictive approach based on the working of the visual buffer is a plausible theory that is not already obviously falsified. This will provide a platform to analyse how it might be used offline in visual imagery in the next chapter and how it is linked more generally to perceptual states in the second half of the thesis (Chapters 5-7).

More specifically, in this chapter I intend do the following four main things. **Firstly**, I will introduce and discuss Depictivism in detail and explain how it is now construed under the Kosslyn Model. This will involve a detailed introduction to what a hybrid depictive representation is actually thought to be. **Secondly**, I will defend this modern version of Depictivism against some initial objections to it. This will involve explaining how topographic arrays in the visual buffer give us a good way to explain how raw visual data can be processed in the early visual cortex. **Thirdly**, I will address a specific issue that certain researchers (e.g. Pylyshyn 2002, 2003) have with Depictivism, in terms of questioning whether these topographic arrays are actually always used in imagery. This is where the idea that the visual buffer is taken offline and re-used in visual imagery is first explored and defended. **Fourthly**, I will analyse why there may still be resistance to the Depictivist approach and suggest some problems that alternative theories may have. By the end of the chapter we should have a fairly solid grasp on what it means to adopt a hybrid depictive approach based on the workings of the visual buffer.

On a terminological note, in what follows, and merely for ease of reference, I will refer to Kosslyn, Thompson and Ganis' (2006) book: '*The Case for Mental Imagery*', simply as "Kosslyn". I will distinguish other work by Kosslyn and his collaborators by referring to it explicitly with a publishing date. I will do a similar thing with Pylyshyn's (2003): '*Seeing and Visualizing: Its Not What You Think*'; and Tye's (1991): '*The Imagery Debate*', and simply refer to them as "Pylyshyn" or "Tye" unless qualified otherwise. I take the Kosslyn and Pylyshyn books to represent a comprehensive explanation of their latest thinking on the Imagery Debate and hence to include and update much of their previous work. Which is why I mainly focus on these texts in what follows. Including Tye's slightly older work where relevant will help elucidate the debate and will also be useful when we discuss his version of Representationalism in Chapter 6.

3.1 – Depictivism and the Imagery Debate

The dialectic in the Imagery Debate over the last 30 years has mainly been between Kosslyn (Depictivism) and Pylyshyn (Propositionalism). As mentioned above, I will argue for a hybrid account that aims to take the best parts of each theory. This is the approach that is currently favoured by Kosslyn and I will suggest later that an early *version* of this is possibly also introduced and defended by Tye. Kosslyn's position is now fairly modest when compared to his original stance, and he argues only that depictive representations are *at least sometimes used*, even if propositional representations are also still widely used elsewhere. So not only is it a hybrid approach but it also allows for mixed representational styles to be used in a task appropriate way in the brain. This seems to fit broadly speaking with the way Fodor (1975) saw things when he first argued for Propositionalism (see opening quote). In other words, even Fodor, the father of Propositionalism didn't initially intend to argue that representations in the brain were exclusively propositional. This contrasts with Pylyshyn's approach which denies that depictive representations are necessary at all. I will analyse Depictivism in detail in the next sub-section and this will require only the very basic introduction to Propositionalism which is given next.

We can treat propositional representations as 'linguistically' encoded descriptions or propositions of what is represented. For example, a ball sitting on a box could be propositionally encoded as "ON (BALL, BOX)", where the relations of the two entities in the example (the ball and box) are described by the relation ON and some rule that the one listed first is on top of the other. It is given this formal notation to emphasise that this is not necessarily described in the mind in a natural language like English, but instead may take on a form that can be loosely described as "a language of thought"[44]. I will discuss Pylyshyn's arguments against the Depictivist further in Section 3.3.

Before we launch into the details of the Imagery Debate, it is worth noting that the imagination-as-simulation hypothesis is not crucially affected by the eventual resolution of this debate (if such a thing is even possible). This is because irrespective of the underlying nature of the visual representations that any theory suggests, the simulationist can still potentially talk of taking these systems offline to form visual imagery. That is even if

[44] C.f. Fodor's (1975) arguments for 'mentalese' and researchers that followed in those footsteps – c.f. Aydede (2004) for a good survey of this). The characterisation above is based on Kosslyn (pg. 10), but is not crucial for what follows, since I will mainly be defending Depictivism and not attacking Propositionalism.

Propositionalism were to eventually be shown to be true, it would still be possible that the simulationist could provide a theory that is compatible with this type of theory and it may just be that we would just be taking a different kind of system offline. As Currie (1995a: 38) indicates: ‘The Simulation of Vision Hypothesis is consistent with both theories.’ Hence in what follows it is not really *necessary* for my purposes to fully defend one theory or to disprove the other.

Having said that, what I will try to do is explain the parts of the debate that are relevant to discussing the nature of STIm and which give us a plausible mechanism by which visual imagery occurs. To that end I do tend to favour explanations given by Kosslyn, because in my opinion they are very detailed and have a significant explanatory advantage over other theories purporting to explain visual imagery. I also think that the latest version of the Kosslyn Model seems to have dealt with most of the current criticisms levelled against it at present and hence remains a plausible and compelling theory. Explaining how it does this is one the main purposes of this chapter and this will hopefully justify adopting it as well as explaining some of the details that will be crucial to the arguments that follow. So we can now look at the Depictivist approach to visual imagery in more detail.

3.1.1 – Depictivism and functional depictions – introduction and definitions

In the early days of the Imagery Debate the depictive theory of representation in visual imagery was referred to as ‘the Picture Theory’, and the idea that there were literally ‘pictures in the head’ was fairly well refuted (c.f. Block 1983). This was mainly on the grounds that taken literally it created other insurmountable problems. For example: there is no evidence for these internal ‘objects’; there is no light to see these internal pictures; there are no internal hands to manipulate these objects; and there are problems of regress in terms of an internal ‘minds eye’ to see these pictures. Hence the more modern versions usually talk of quasi-pictures that are picture-like, and share *some* properties with pictures, but are not literally pictures in the head. Unfortunately this 'quasi-picture' terminology still runs the risk of being confused with early picture theories and so I will refer exclusively to 'Depictive Theories' or D*epictivism* to make it clear that this is an importantly different proposal[45].

The reason that Depictivism is related but crucially different to Pictorialism is that, although they both appeal to a depictivist style of representation, Depictivism does not suggest that anyone literally looks at these representations in the same way that a picture would be looked at. So even though a picture can depictively represent a 3D visual scene by representing it depictively on a canvas in 2D space, which someone can then observe, in modern day Depictivism we do not suggest that a spatially laid out representation in the brain depicts by way of someone looking at it at all. Rather it depicts by the way that it is stored spatially and the way that it is accessed. However this second non-pictorial kind of depiction needs some introduction to make it clear what we mean.

To that end we need to be clear about what Depictivism entails and Kosslyn (pg.11-13) provides the following definition of what counts as a depictive representation according his school of thought:

[45] See for example Kosslyn (pg. 38) for an explicit statement on this point (e.g. 'To have a picture in one's head would be very uncomfortable'). And Pylyshyn (2003) is still sometimes guilty of setting up a straw man in this respect when criticising Kosslyn's work. However there may be a reason for this as discussed in Section 3.1.3.

Definition – Depictive Representation

Syntax:

1) Symbols belong to two form classes: points and empty space
2) The points can vary in size, intensity and colour
3) The points can be arranged so tightly as to produce continuous variation, or so sparsely as to be distinct (like dots in a comic strip)
4) The rules of combining symbols require only that points be placed in spatial relation to one another

Semantics:

The association between a representation and what it stands for is not arbitrary, rather, depictions "resemble" the represented object or objects. That is:

a) Each portion of the representation must correspond to a visible portion of the actual object or objects
b) The represented "distances" among the portions of the representation must correspond to the distances among the corresponding portions of the actual object (as they appear from a particular point of view)

Table 3.1 – Kosslyn et al's (2006: 13) definition of depictive representations

I intend to adopt, utilise and develop this definition in what follows. As a way of explaining the above points lets take a pointillist painting of a person's face as an example. This is like the comic strip example described above where coloured dots are used to fill in the areas of the painting:

Figure 3.1 – Examples of Pointillist paintings (except for face on bottom left)

With these paintings the points can either be made out individually or they will merge to make a continuous area of colour depending on how far away you view them. We can now ask how would this fulfil the depictive criteria given in Table 3.1? It is fairly obvious how the *syntax* is equivalent since the paintings are mainly composed of spots of colour. Classically the points will remain a constant size across the painting and vary only in colour and intensity. However there is no reason why they might not get larger to the edges of the painting and be very small and detailed in the centre. This way of describing it mimics what is thought about vision, where finer detail is processed at the centre of our focus of attention (see later). The *semantics* part of Table 3.1 is to do with preserving the distance relations in the representation that in some way resembles those in the original. So this could literally be having the painting the same size as the original face and keeping all the facial features in the same relative positions (see for example the two faces on the left of Fig. 3.1). Or it could be an expanded version that is ten times the size, with all the facial features scaled and placed accordingly. The condition of keeping the distance relations faithful is compatible with a similar condition expressed by Tye (pg.36) in a bit more detail and is captured in the slogan: *depictive representations use space to represent space*. This ensures that, for example, the time taken to scan between two features in the original will be *proportional* to the time taken to scan between them in the depiction. Obviously they do not have to be *exactly equal* in time, as there may be a difference in scale to adjust for.

To develop this concept further, Kosslyn (pg.12) introduces *functional depictions* as another way that something could depictively represent another thing. Here he uses an analogy of images stored in computer memory in an array that could still *depictively represent the geometric relations* of the original. An example of this is how bitmap images are stored on a computer or are presented on a computer screen. Here each pixel representing a part of the original picture is assigned a coordinate scheme in a cell of an array (e.g. top left = [1,1], bottom right = [100,100]). These cells can also be allocated certain other parameters, such as its represented colour. Users of computer illustration software will be familiar with a colour on the software's colour palette being, not only visually represented as colours on the screen, but also presented nearby as an alphanumeric proportion of, for example, red, green or blue[46]. The idea being that you can get any colour on the palette by 'mixing' different amounts of red, green and blue in an analogous way to mixing paint. Using this method you can fill in an area with colour by representing it visually on a screen and/or storing it in a more propositional style format on the hard drive. For example, a line of red on the top left of the screen could be represented like this: pixels or array cells [1,1 to 1,10] = RED (R-255, G-0, B-0). Or another way of putting it is that array cells [1,1 to 1,10] contain the information that, when they are accessed appropriately, should be visually represented by a row of pixels at the top left hand corner of the image that are coloured red. Hence it seems possible to represent the same

[46] The exact colours suggested here are not important and green could be substituted for yellow, for example. Alternatively they can also be presented in terms of a hue, saturation, brightness (HSB) scale. The point is that these colours can be represented alphanumerically in a way that is more similar to a propositional format. For example, in the illustration software on my computer: RED is presented as (R-255, G-0, B-0); BLUE as (R-0, G-0, B-255); BLACK as (R-0, G-0, B-0); and WHITE as (R-255, G-255, B-255), and so on.

information in an obviously depictive format on the screen, or in a more alphanumerical proposition-like format on the hard drive.

The above will become useful in what follows, but for now let's return to analyse Kosslyn's claim that these functional depictions can still preserve the *spatial relations* of the represented object. In most computers a bitmap image can be stored in a distributed way across a hard drive, so that adjacent cells in the array need not literally be stored anywhere near each other physically. However, although the physical locations of the array cells on the hard drive are assigned arbitrarily, they can still be accessed sequentially, *as if* they were located next to each other. Just so long as the process that accesses them retains the information of which cell is stored where and they are indexed and accessed appropriately. Hence it would still take a *proportionate* amount of time to, for example: scan across the top of a computer array, cell-by-cell, looking for the equivalent of a 'red dot'; as it would to check, looking from left to right, at each dot on the top edge of a pointillist painting or equivalent screen image. That is, each entry would need to be examined for its contents in turn to see if it matched the target criteria of representing a 'red' cell or pixel. And it would take proportionally longer for each process the further the 'red' cell is from the top left hand corner of the image. In principle an equivalent story can be told for other more complex scanning and rotation searches e.g. diagonal scanning and rotating to match a target. The point is that the array and the processes that work on it can be designed to treat the computer-based representation as if it represented space with 'virtual space' within the hard drive. In other words, although the array is not spread out in literal 2D space, the array can be accessed and manipulated to mimic what would happen if it were.

It is worth noting that for this to happen, these arrays of stored information have to be organised and treated as entities that represent functional space. It is not enough for these arrays to simply exist as indexed pieces of information to make them depictive. This is because these cells could be accessed in many different ways that wouldn't necessarily represent their relations in real space. Therefore the order in which they are filled and the way they are manipulated is as important as the information that each cell contains. If these criteria are respected, then these arrays and represented virtual "distances" can fulfil the syntax and semantics requirements listed above, and this is what qualifies them as *functionally depictive representations* according to Kosslyn's definition. If this is the case then these arrays can be made to function as if they are actually depicting real space and hence qualify as functional depictions. In slogan form: they represent real space with functional (or virtual) space. This analogy can now be used to analyse the equivalent of this that is thought to occur in the visual buffer under the Kosslyn Model. So now we have an idea of what Kosslyn means by Depictivism and a functionally depictive representation, let's look at how he applies this to vision and visual imagery.

3.1.2 - The visual buffer and hybrid depictive representations

In Section 1.3, I first introduced the idea of a visual buffer and offered some quotes from Kosslyn to back this distinction up (see Kosslyn pgs. 18-19 & 135-6). In this section I will add to this initial introduction and describe how the visual buffer is thought to work in much more detail. To recap, according to Kosslyn, vision and visual imagery are primarily processed in what he calls the *visual buffer*. In terms of *vision* the visual buffer cells are organised into arrays, which contain information about the visual scene in front of you. This information is based on wavelength detections (colour) and binocular disparities detected at the eyes (distance). This information is sent from detectors in the eyes, to the back of the brain where the visual buffer is located, as a sub-area in the early visual cortex.

It is generally accepted that different aspects of visual experience, such as: colour, distance, edge detection and movement, are processed by different localised areas of the visual cortex. This means that at this stage at least, it must operate in a parallel-distributed fashion, rather than in a serial or sequential manner. Hence each functional cell is actually made up of certain linked areas from these different topographically organised parts of the visual cortex (and see Fig 1.4 for an illustration of this). Kosslyn explains the nature of the visual buffer further in the quotes below:

> ‘We group the topographically organised areas in the occipital lobe [visual cortex] into a single functional structure, which we refer to as the *visual buffer*. Patterns of activation within the visual buffer depict shapes, according to the definition of depiction offered here [see above]. The visual buffer, in essence, is [like] the canvas upon which images are painted; it is the medium that supports depictive representations.’ Kosslyn et al (2006: 18 – my inserts in square brackets)

> ‘In spite of their coding nondepictive information, these hybrid representations cannot be reduced to propositional representations. Crucially, they use space (literally, on the cortex) to represent space in the world. The fact that each point codes additional information does not obviate its role in depicting shape.’ Kosslyn et al (2006: 18-19)

There are several things to point out from the above quotes. First of all, in the top quote, it is regrettable that the ‘canvas and paint’ terminology is used without more care to qualify this as *merely* an analogy. As mentioned above, much damage still occurs because critics use this kind of talk as indicating obvious problems with the *pictorial* approach. Hence I have inserted the word “like” in square brackets in the quote above, but it’s very possible that this line will be quoted out of context to show there are still literal pictorial assumptions at the heart of this approach. This kind of talk works merely as an *analogy* because, as with placing paint on a certain physical point on the canvas, filling one of the visual cells of an array with a representation of colour, would arguably have the same ‘virtual’ depictive effect. So warnings aside, the top quote does make explicit the functional role of the visual buffer as the medium that supports depictive representations.

The first quote also backs up the functional description of the visual buffer as described in the first quote on the visual buffer in Section 1.3. To briefly re-iterate those points here: the visual buffer is limited to topographically organised areas, and by extrapolation from the monkey related empirical data, there are probably at least sixteen of these areas in the human visual cortex. It also seems that damage to localised parts of these areas directly relates to impaired vision in corresponding modalities (e.g. colour). I have also explained in Section 1.3, how damage that is further apart in the visual buffer can lead to blind spots in the visual field that are farther apart (c.f. Figure 1.4), hence this also supports this functional definition and structural characterisation.

Related to this, the second quote above is of the most interest for the rest of the discussion, because it represents an important concession on the part of Kosslyn. Here we are introduced to the position he now favours in terms of hybrid depictions. This position suggests that not only do the cells of the array represent depictive information, but also more abstract and *propositional* data is also appended to them. Referring back to Kosslyn’s definition in Table 1.3, we can see that each cell can vary in the following ways and still remain depictive: the points can be empty or filled; vary in size, intensity and colour; and vary in density. So, as an example, for an area of the visual buffer that only processes colour, it is feasible that each cell could remain depictive since a functional representation of colour in a cell, not only indicates what colour the cell is filled with, but also that it is filled. This seems to be equivalent to a coloured dot on a pointillist painting in certain crucial ways: the colour of the dot not only represents a coloured area on the original face, but also represents the fact that there is something there to be represented. This clarification will be useful in what follows.

This version of Depictivism is now very similar to Tye's (1991: 91) approach, which seems at least sympathetic to Kosslyn's later hybridism, but different in one relatively trivial respect. This is because Tye limits the depictive part to be a simple statement of whether a cell is filled or not. This relates to a condition of his that: 'every part of the representation represents a part of the object' (pg.36). He refers to this aspect as an O-symbol and other symbols representing, for example: colour, brightness and depth, are also appended to the contents of this cell. However I don't really see the need for this O-symbol, whose sole purpose is to indicate if the cell is full or not, because surely simply the presence or absence of *any symbol at all* does the same job on its own. By analogy when you look at a colour spot on a pointillist painting there are not two different things you see, but one thing that represents at least two things. For example, a red point indicates: a) that the location is full; and b) that it is filled by red colour. So I think we can get rid of requirement for an extra O-symbol in Tye's account without loosing any relevant functionality. The cells are either filled with symbols representing something, or they are empty and are equivalent to an empty space. A symbolic entry representing *at least* colour still represents a part of the object, hence meeting Tye's condition in a more parsimonious way. Hence while I think we can adopt much of Tye's early work in this area in this respect, we can perhaps do away with his O-symbol requirement.

In terms of the potential for the cells to vary in density and to represent depth, consider the quote below:

> 'The high-resolution central (foveal) portions of these areas [in the visual buffer] represent the central portions of the visual field, where as low-resolution, peripheral portions of these areas represent more peripheral portions of the visual

> field. Space on the cortex is literally used to represent space in the world (more precisely, space on the cortex is literally used to represent a planar projection of space in the world, because the cortical areas are two-dimensional; information about depth is represented in a different manner…).' Kosslyn (2006: 103)

This indicates that central areas in the visual buffer have a higher density of cells representing what's in the world at the corresponding centre of your focus of attention. Analogously these areas are depicted with more densely packed points or with 'finer brush strokes'. Conversely areas to the periphery of your main attention have fewer resources devoted to them since they are almost by definition of less interest to you. Analogously these areas are depicted with less densely packed points or 'broader brush strokes'. This is basically what you would expect in a predator with forward facing eyes, who needs to focus in detail on its prey and be less sensitive to the danger all around them. So this density variation does not violate Kosslyn's conditions for depictive representation.

The condition for the depictive representation of depth is more complex and interesting. Light is sensed in the eyes on a 2D surface at the back of the eye cavity, which responds to the light distribution in a topographic or *retinotopic* way. This is purely a function of the geometric way that light converges on the retina through our cornea and the distribution of sensors (cones and rods) in the retina that detect that light. Hence at any one point on the retina, 3D information about the world is not sensed in a single eye, but only the 2D distribution of wavelengths incident upon it. However depth information can be recovered further upstream, by looking at differences between the light paths to each eye (binocular disparities), and using this to calculate the distance to the object. Kosslyn suggests in the quotes above that depth (or distance to object) is therefore represented in a non-depictive manner and this would appear as an appendage to each 2D cell in a topographic array in the visual buffer. Hence any array that contains at least depth and colour information must be a hybrid depictive representation in the following way: the cells that are filled and are depictive of a 2D flattened scene in front of you (like a painting of a landscape); and some of the information they are filled with is non-depictive in that it describes certain things one might find in those related areas, like depth and colour representations. Hence we have a mixture of a depictively organised array, which contains propositional information in each of its cells when filled.

This explains what Kosslyn means by hybrid depictive representations and how they are filled in the visual buffer. This seems to me to be a plausible and intuitive way of describing how all the rich information collected at the eyes is represented and processed efficiently in the early visual cortex. However at this point the sceptical reader could still question whether the way these arrays are topographically arranged is just a coincidence, or whether they are actually crucial to how the visual buffer functions as a whole. Hence we now need to look at evidence that supports this further claim that these topographic areas are functional.

3.1.3 – Evidence for topographic arrays and their function

The first evidence usually referred to that suggests these topographic areas exist is from Tootell et al (1982) and this is illustrated in the figure below:

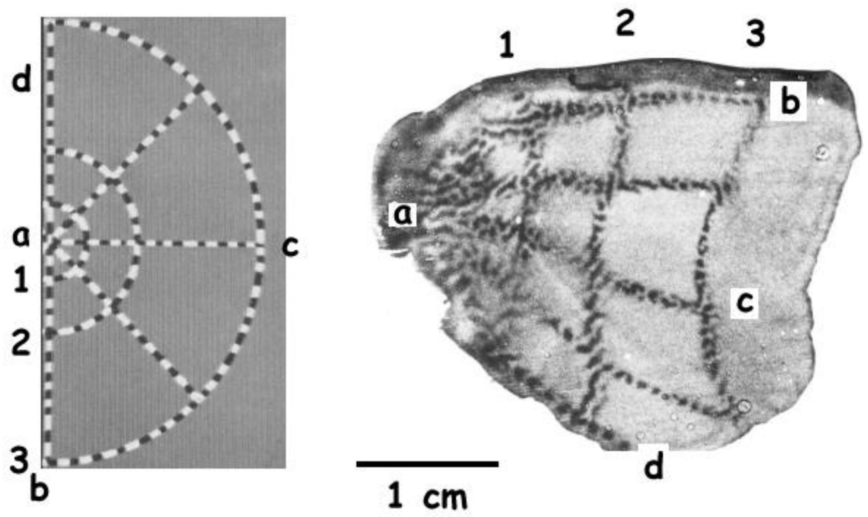


Figure 3.2 – Example of stimulus flashing light pattern (left) and the preserved visual cortex reaction in macaques (right) as indicated by higher activity in darker areas – adapted from Tootell et al (1982)

What the above shows is areas of the macaque visual cortex (right) firing in a topographically matching way to a set of flashing lights set out in a wagon wheel formation (left). Before the animal is sacrificed it is fed isotopic sugars that can be traced to areas of higher activation (darker areas) that respond to lit areas on the wheel. Areas that respond to the inter-light spaces show lower activation and hence require less energy and use less sugar (lighter areas). The visual cortex is then sliced and measured to yield the right hand part of Figure 3.2. It is quite clear that the spatial distribution of the lights is represented in an equivalent spatial manner in the cortex of the monkey. This is good evidence that the visual cortex is topographically organised in monkeys, and claimed by Kosslyn to show that it depictively represents the external world.

Kosslyn (pgs. 101-107) also provides references and a detailed analysis of other experiments that suggest there are multiple topographically organised areas in the human visual cortex. Among those are ones that show that the distance between two areas of the visual buffer that are damaged, directly relates to how far apart experienced scotomas (blind areas) will be. Hence showing there is a spatial link between what is processed in these areas and what is experienced in the visual field (c.f. Section 1.3 and Figure 1.4). The details of this need not concern us here, since both sides of the debate can accept that these areas exist, but obviously attribute a different amount of significance to these results[47].

The key concern to the imagery debate is whether these topographically organised areas really qualify as depictive representations. The first point to note with regards to this is to ask why these areas are literally topographically organised, rather then just functional representations distributed over the visual cortex; like the distributed storing of the bitmap image on a hard drive. There is no a priori reason why information from adjacent sensors in the retina need be stored in adjacent areas of the visual cortex. Related to this, a lack of evidence for topographically organised areas would not necessarily be evidence against *functional* depictive representations (c.f. Kosslyn pg. 50). So what is achieved by copying retinotopic sensory arrays from the eyes into various topographic arrays of the early visual cortex?

Kosslyn (pg.17 & footnote 5 on pg. 23) offers this suggestion as to why we do actually possess and use these topographically organised areas: it has evolved as the most efficient way to process visual information. He uses the example of edge detection to illustrate this point. If an area that is reflecting a lot of light (an object) is next to an area that isn't (the background), it is useful to detect this change by emphasising the difference between them and potentially detecting this as an edge of an object. One way of doing this is for neurons that are representing the bright area to inhibit areas that are representing the dull area and hence accentuate the difference. For these areas to interact efficiently it is better for them to be nearer each other physically in the cortex. This is because sending signals down long neuronal paths has an energy cost that is avoided by keeping things that need to be

[47] Cf. Pylyshyn (2002: 174-180) and Pylyshyn (2003: Chpt.7) for a negative critique of the significance of these results.

regularly processed together, closer together (and recall Hurley's slogan from Chapter 2: what fires together, wires together). Since this is going to occur across the visual field, it makes sense efficiency-wise to keep functionally closer areas actually closer to each other physically. Hence topographic mapping is potentially the result of efficiency constraints imposed during the evolution of the visual system. So while adjacent functional cells do not essentially have to be near each other it makes evolutionary sense that they are, as it will speed up processing and save valuable energy better spent on other survival tasks.

This seems like a plausible explanation and can be further illustrated by a simplified worked example of edge detection below. Introducing this also starts to link this with Marr's (1982) seminal work on how visual processing could occur utilising step-by-step algorithmic computations. It seems that Marr's theory still remains one of the best attempts to explain this process even today. So the ability to match up with it can be seen as a considerable theoretical bonus. This is a sentiment that is echoed in Tye (pg.93) and he also runs through a similar simple edge detection example as the one below (pg.79).

In the example below greyscale data is represented on a scale of zero to nine, with zero being black and nine being white. The object in this example is a dark grey square (greyscale 2) that is detected against light background (greyscale 7). The raw data is organised in an array from top left [1,1] to the bottom right [7,7], and each cell is filled with a measure of the greyscale corresponding to that area on the object (in this example either a 7 or 2). The first thing that happens to the raw data on the left is that adjacent vertical columns are subtracted from each other, which reveals the vertical edges represented by 5 and –5 in the processed array on the top right. Then the same is done on the original array for the horizontal rows, with the same result in terms of the horizontal edges (middle right array). Then the two processed arrays are summed to give the bottom array on the right and we get a rough idea of where the edges of the square are by linking up and extrapolating from the 5 and –5 entries[48]. So this represents a very simplified example of how adjacent cells in arrays could be processed to yield a kind of rough edge detection.

Topographic representation of a square

7	7	7	7	7	7	7
7	7	7	7	7	7	7
7	7	2	2	2	7	7
7	7	2	2	2	7	7
7	7	2	2	2	7	7
7	7	7	7	7	7	7
7	7	7	7	7	7	7

Horizontal (column) subtraction

0	0	0	0	0	0	/
0	0	0	0	0	0	/
0	5	0	0	-5	0	/
0	5	0	0	-5	0	/
0	5	0	0	-5	0	/
0	0	0	0	0	0	/
0	0	0	0	0	0	/

Vertical (row) subtraction

0	0	0	0	0	0	0
0	0	5	5	5	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	-5	-5	-5	0	0
0	0	0	0	0	0	0
/	/	/	/	/	/	/

Addition of processed arrays

0	0	0	0	0	0
0	0	5	5	5	0
0	5	0	0	-5	0
0	5	0	0	-5	0
0	5	-5	-5	-10	0
0	0	0	0	0	0

Shows rough edge detection of square

Figure 3.3 – Example of a number filled array spatially depicting a darker square on a grey background and how this can be used functionally to detect edges

The above gives a very simple illustration of how topographically organised areas could interact with each other more efficiently if they were located nearer to each other in the

[48] I accept that there is a zero and –10 in this example but this is something that could be tidied up using more sophisticated algorithms or more steps in the process. And obviously, if desired, this example could be made more realistic by inserting the RGB [0-255] colour palette values discussed earlier, but I have tried to keep things as simple as possible to make my basic points.

brain. To return to linking this with Marr's theory of visual processing, Marr would suggest that each cell in this 2D array is also appended with: depth information; orientation of the surface relative to the viewer; and other 3D structural information in symbolic form e.g. whether it's an edge or ridge. He calls this a '2 ½ D sketch' since while it is still structured as a 2D array, it does contain the data necessary to convert it to a 3D image (c.f. Marr 1982: 129). Tye (pg. 83) compares this with Kosslyn's definition of quasi-pictures at the time (i.e. 1991) and suggests that Kosslyn makes no allowance for these appended symbols representing depth and orientation etc. However we can see from the above that Kosslyn's recent adoption of a hybrid depictive approach does now make allowance for it. And this is one way that Tye and Kosslyn's hybrid approaches are now perhaps very similar and are both compatible with Marr's work in the area.

Kosslyn (pg.107) also points out that these topographic arrays are not just copies of the retina but are doing the ground work of processing the data in a step-by-step fashion as Marr's theory predicts. Each stage uses processed data from the previous one and as it flows upstream it can become less and less topographically organised. For example, representations in the temporal cortex (at the end of the ventral visual stream – see Appendix B) may be only loosely topographically organised, but those in the parietal cortex (at the end of the dorsal visual stream) retain some topographically organised areas[49]. So Kosslyn's definition of the 'visual buffer' as the functional combination of the topographically organised areas in the visual cortex is equivalent to identifying it with the areas that are doing the raw processing work based on hybrid depictive representations. Perhaps then a good way of representing the contents of each cell in the array in a hybrid fashion is as follows:

---

[49] I will explain the terms dorsal and ventral stream in Chapter 4 (but see also Appendix B). And note we will see later that the topographic areas in the parietal cortex might be processing spatial information, rather than detailed depictive information.

Array Cell (X, Y) = [colour (RGB: P, Q, R), depth Z, surface orientation B, edge/ridge D, etc]

In the annotation above each entry (e.g. depth, colour) could potentially be stored and processed in a separate topographic array but would be functionally linked to each other by their cell designation. That is distributed cells processing only one aspect of vision would be functionally linked so that the combination forms a virtual cell that contains all the information annotated above. The last two entries (i.e. surface orientation and edge/ridge), incorporate information, which Marr suggests the 2½ D sketch would also have, as described above. I will utilise this notation and expand on it in what follows, although I will mainly focus on colour and depth and generally ignore the last two entries to keep things simpler. But note they can be added again at any point should any future study in this field become more focused on this aspect.

Given the above, what is still missing from everyone's account is how this 2 ½ D sketch is converted to full 3D perception of a visual scene, or a perception-like phenomenal image containing apparent shapes and colours. This aspect will be discussed further in the second half of the thesis. But note that this comes close to requiring a solution to the hard problem of consciousness, which I will not cover in this thesis (see Chapter 7). What we can say here is that we now have a fairly detailed description of how the sub-personal mechanisms potentially sub-serving vision and visual imagery may work up to this point according to the Kosslyn Model. We can now look at Pylyshyn's first main objections to this kind of approach.

3.2 – Potential problems with functional depictivism

In this section I will ignore much of Pylyshyn's criticisms that seem to be directed at older versions of pictorial theories of visual imagery. Instead I will focus on what his objection to Kosslyn's modern Depictivism and particularly the *functional hybrid depictivism* described above. Before doing that we need to look at a part of the debate that makes explicit how a Propositionalist would deal with some of the experimental results that seem at first glance to support Depictivism. At the base of this explanation is an option for the Propositionalist to appeal by to 'padded lists', which contain dummy nodes representing empty space but in a non-depictive way. I will try to show that these padded lists may be close to describing a hybrid functionally depictive array anyway and why there may be some commonality at the core of these approaches.

To discuss this point we can look at the now classic debate[50] concerning empirical studies involving map scanning and mental rotation times that occurred in the 70's; and note I will focus mainly on the map scanning results here. In these experiments subjects were asked to memorise a map of an island with various landmarks on it at different distances from each other. They were then asked to *imagine* scanning the map from one object to another and the time it took them to scan between different objects was found to be proportional to actual distance between the objects on the map. This seems to indicate that imagery occurs in a depictive way, because time to search a functionally depictive array could explain why it is proportional to actually scanning a map with the eyes. However it turns out that it is possible that a propositional *description* of the same map that describes the whole map including empty spaces can also account for this phenomenon. This is because the time taken to access such a propositional padded list would vary proportionally with distance, since greater distances would have more padded entries (or equivalently more nodes to go through).

[50] C.f. Kosslyn et al (1978), Shepard and Metzler (1971); and Block (1981) for a collection of papers in this areas.

Another way of explaining these results is to suggest that you are using your tacit knowledge of how real visual scanning works to mimic this effect when scanning a visual imagery. That is you may be sub-consciously taking time to 'examine' even empty spaces or propositional descriptions along the way so that your scanning time in mental imagery matches that in real world scanning. However there are now example of these effects that cant be due to tacit knowledge[51] and sine this aspect of the debate is less useful to our concerns here, I will not elaborate on these any further. I mention them merely to make the reader aware of another aspect of the debate.

To try and assess the claim that a padded propositional lists like this is *always* used, Kosslyn et al (1978) changed the design of the experiment so that it was possible for subjects to simply decide whether a named object was located on the island. In this situation they reasoned no scanning would be required as its possible that the subjects could just memorise a list of the names of objects on the island. And this is exactly what they found; the time to decide if the named object was on the island was not proportional to distances between them on the map. So if a propositional list is *always* used according to the propositionalist, why do we see a difference in these results? The propositionalist response to this is that there is no limitation on what kind of list a person can use: it could be a padded list or just a list of names. And one can simply access a list that represents the data in the most efficient way to complete the task. So it is possible that there are *at least* two lists available to the subject in this task and the propositionalist potentially has the resources to invoke other kinds of lists with different properties to explain any new data. This gives one an initial sense of how ad hoc the propositionalist position might turn out to be. Consider the following remark from Currie on this matter, where he is referring to similar results observed in mental rotation experiments:

[51] See Kosslyn (Chapter 3) for a good review of these.

> 'She [the propositionalist] can neither predict in advance of the observations that these response times will be observed, nor offer any plausible account of why they take the form they do. She is like the Ptolemaic astronomer who cannot predict the observed features of planetary motion in advance, and whose activity is confined to bolting on more epicycles every time a new fact is discovered. That, at least, is the pictorialist's claim.' Currie (1995a: 39)

To accentuate this seemingly ad hoc 'proliferation of lists' problem we can even imagine that there would potentially need to be a padded list to cater for every possible travel path at any angle form any point on the map to every other point (i.e. a full 360 degrees for every point). This would be needed so as to make sure all the *possible* scanning tasks in any direction would be able to be mimicked. One way out of this might be to suggest that a cell-by-cell padded list description of the whole map is created and stored in the brain, but this now sounds suspiciously like the 2D functional depictions described earlier. We will look at how closely these match up next, but first we can note the general acceptance by Kosslyn (pg.30) of a point made by Anderson (1978): that as long as we just have behavioural data to go on, then given a preferred underlying representational format, it will always be possible to adjust the processes that work on it to explain the results. In other words the propositionalist can always change the structural set up so as to mimic the properties explained by the depictivist account. Hence these kinds of experiments will never be able to decide between the two camps apart from on grounds of parsimony, and this will always only ever be suggestive evidence rather than provide anything more compelling. This explains why the debate has now moved on from this kind of discussion.

So returning to the analysis of how the propositionalist's padded lists may start to resemble the Depictivist's functional arrays, Pylyshyn gives the following general objection to appealing to functional depictions in the first place (and note he uses the term matrix here to refer to an array):

> ‘The problem with the functional space proposal is that functional spaces do not have any particular functions intrinsically. Being functional, they are not subject to any natural laws and therefore can be assumed to have whatever properties are needed to account for the experimental data. Because a functional space has no intrinsic properties, any properties it has are stipulated or extrinsically assumed, and so it can accommodate any findings whatever. Or, to put it more positively, if the extrinsic assumptions provide any explanatory advantage at all, they can be equally well adjoined to a theory that assumes *any* form of representation, not just a pictorial one [i.e. a propositional one]’. Pylyshyn (2003: 360)

> ‘Notice, moreover, that both [Depictivism and Propositionalism] become completely principled if they are taken to be simulations of a real spatial display… you might wonder why the matrix feels more natural than other ways of simulating space… [this may be] because we are used to thinking of and displaying matrices as two-dimensional tables… because of this, it is natural to interpret a matrix as a model of real space and therefore it is easy to make the slip between thinking of it on one hand merely as “functional space”, and thinking of it on the other hand, as a stand-in for (or a simulation of) real space – a slip we encounter over and over again… But we must recognise that in this case we *are assuming that images are written on a literal spatial medium,* which we happen to be simulating by a matrix (for reasons of convenience).’ Pylyshyn (2002: 167)[52]

First of all note the language suggesting that if you are simulating “seeing” real space then it is not surprising you get results similar to those found in vision. This links with my indication earlier that even propositionalists can explain certain aspects of visual imagery as mimicking actual vision, and hence this theory may be accessible to an imagination-as-simulation approach. We can also see from the above quote why Pylyshyn assumes that Depictivism must be committed to some more classically Pictorialist assumptions, because if not then he assumes mere ‘functional depictivism’ potentially collapses into, or is at least compatible with, Propositionalism. Which would explain why he feels he can attack

---

[52] Although these quotes are from different sources they are drawn from sections that are almost identically structured, so can tolerate being juxtaposed while still preserving the crucial meanings. Indeed, much of his 2003 book seems to be based on his 2002 BBS article or vice-versa.

Depictivism on its claims that these representations must have some more classical pictorial qualities.

He goes on to add that it is not that the matrix or array approach to representation is wrong, but only that invoking it is neutral with regards to the current debate. The array only behaves in a way that mimics real space if that is the way it is made to work by the processes that access and manipulate it (i.e. extrinsic factors). But this is a feature of how those processes interact with the array and not a property of the array itself. Hence the array taken on its own has no intrinsic spatial properties and is not in this sense functionally depictive: it need not use space to represent space. For instance, we could take a column or row (or even any two points) from anywhere on the array and subtract them from each other, but this would not tell us where a spatially located edge is. There therefore needs to be some further restrictions to say that we can only find an edge from differences in colour intensity between adjacent rows or columns.

We can illustrate Pylyshyn's point further by seeing that even the Tootell et al (1982) evidence, as introduced above in Fig. 3.2, on its own is not conclusive. This is because, although it is true that when looking at the slice of the monkey's visual cortex with the human eye, it does indeed act to depict the original object. Unfortunately this does not necessarily mean that the processes going on in the monkey's brain are also limited by the spatial distribution of these firing neurons. It seems that it's possible for information from very different parts of the array to be accessed without having to respect how they are depictively laid out physically in the visual cortex. Therefore there needs to be some further way of restricting how these arrays are accessed and manipulated to preserve their status as functional depictions.

In reply to this, it seems that Kosslyn has stipulated for this more recently and this is shown explicitly in the quote below:

> 'However… a functional space is sufficient as a depictive form of representation only if the geometric properties of the representation emerge because there are fixed, hard-wired processes that interpret the representation as if it were space; if the processes are not fixed, then the representation is not necessarily a depiction.' Kosslyn (pg.15)

The theoretical addition that Kosslyn therefore makes is that these arrays *cannot* be treated non-spatially. In other words, it is hard wired into how they are processed that adjacent cells in functional space are treated as equivalent to adjacent parts of objects in real space. Hence, although it is possible that they could be accessed in different ways, as it turns out they **can't** be in *at least some cases*. This is because the way these arrays are processed 'spatially' is fixed by the way they are wired together neuronally and how they are accessed. I have qualified the sentence above with 'at least in some cases' because this is all Kosslyn needs to defend his current position. That is, in at least *one* case functionally depictive representations are used, even if propositional representations are used the rest of the time. Of course the Depictivist would want to claim that they are used more pervasively in the visual buffer, but this just represents the minimal claim necessary for Kosslyn's position to go through.

Perhaps we can now ask the following question: when would we expect functionally depictive topographic arrays to be used in perception and would this be a rare or a common occurrence? Well it seems the efficiency claim described above can be applied to many different aspects of distributed visual processing. For example: from tracking movement, to identifying discrete objects, to processing colour. This is because most objects in our ancestral environment are lumped together into discrete areas with fairly homogenous properties (i.e. colour) and they also move in a fairly continuous fashion (i.e. smooth movement). In other words, generally speaking: rocks, trees, rivers and prey, are all fairly locally grouped and show generally uniform properties of colour and shape. And most objects, including other animals, tend to obey the laws of physics and tend to move smoothly (e.g. running and falling with gravity) and follow fairly continuous paths (i.e. they don't teleport). Also when you move around your environment things generally shift continuously past you. To experience this first hand try walking (or imagine walking) through a doorway, or observe how the scene changes as you simply turn your head.

To me it would make sense that, since spatial objects generally follow smooth trajectories, that any mechanism that is trying to track this would form in a way that is most efficient to process this information. This could mean that information like colour and depth in any single array cell is quite likely to transfer with some, usually minor, modification to the adjacent ones. Hence the priming and passing of information between adjacent array cells is more efficiently done if they are also physically located closer together in the visual cortex. This can then be used to help interpret and process the raw information coming from the eyes and potentially save on processing effort and improve reaction times and avoid making errors. Hence I would suggest we would expect to see this trick used quite often in mechanisms involved in object tracking (vision) and object manipulation (motor actions). This will probably also be mirrored in imagery, where we are predicting imaginary versions of the same kinds of features (e.g. imagining fairly homogenous objects in shapes and colours which are moving fairly smoothly in the environment).

Perhaps we should now ask, why don't we think like this in terms of a computer array? Well there is one important disanalogy in this comparison: there are effectively no power limitations on a computer. Hence any small power savings such as this would be pointless to develop since it is relatively uncostly to transfer data from adjacent array cells from physically distant parts of the hard drive. If power resources were at more of a premium in computing, then this might be seen a good way of designing-in some novel energy saving processes. Conversely in the human mind, the brain is one of the most power hungry organs, using up to eight times the energy of equivalent cells in other organs (cf. Dunbar 2000: 246). Hence any minute power saving would probably be positively selected for, and this could include arranging adjacent cells physically near each other so that often transferred information can take as short a route as possible.

This effect may also be accentuated by the parallel-distributed processing style of neural networks in the brain. It may be that the job of tracking and processing real world objects is best done in topographic arrays as this suits the way neural networks inhibit and excite nearby neurons. There also seems little point in loosing 2D spatial information available indirectly from the retinotopically organised projections to the visual cortex, and it makes sense to at least use this as an *initial* way of sorting out the raw data. Hence giving further reason for the visual buffer to organise itself and utilise topographic arrays that operate to functionally depict real space. It seems that when these are viewed indirectly in operation by a third party (e.g. in a brain scan), they also seem to actually depict 3D space on a 2D medium. Hence explaining the mounting empirical evidence that these topographic arrays exist and are functional in visual processing (see Section 7.2 for a very modern example of this from Kamitani et al 2008).

It seems then that there may be a number of reasons why light reflected from external objects would be processed in a topographic manner in the visual cortex. It may therefore be enough to prove Kosslyn’s point by suggesting that, although it is not necessary for these arrays to use space to represent space, it turns out it’s quite a good evolutionary strategy to do so. As long as the empirical data continues to support the existence of these areas then this seems like a good explanation of why they are there. And further to this, it seems that any theory that invokes array-like nested lists to represent space, is quite likely to do this also (e.g. padded lists). Why bother making an array unless you are going to use it in an array-like way and why not also lay it out in the cortex in the most efficient way possible. In light of the above, even the Propositionalist would predict these topographic areas might be used occasionally to organise nested lists arrays in a way that mimics a hybrid depictive array.

Related to this, it seems that some of the work that inspired Pylyshyn's Propositionalism in the first place, used 'nested lists' to solve spatial puzzles based on a programming language called LISP. LISP was an early example of how a propositionalist-style programming language could be used to solve problems in a potentially analogous way to humans. Thomas (2008: Section 2.2) quotes work by Baylor (1972) and Moran (1973) to illustrate this point. In those studies, in order to solve a 'block slicing' puzzle, nested lists were used to represent space: e.g. faces, edges and vertices. Hence the invocation of things like nested lists or arrays, to represent space is a common feature to both sides. I think the main debate now is whether having them laid out in space in topographic arrays in the brain means they have a depictive element, or whether this is just Propositionalism with a physical efficiency constraint laid on top of it. As mentioned earlier, I am not much interested in determining which theory is more likely to be correct, as both are accessible to the simulationist. It seems there is convergence on some points and it may be academic what we call it. The main thing is that we have seen in some detail what the issues are concerning this debate. That said, I do tend to prefer Kosslyn's approach as it has much greater consistency and explanatory potential and for other reasons as described above and in what follows.

It's also worth pointing out that presumably propositional arrays of this detail (i.e. a cell by cell nested list) will be required for every moment of vision, since typically we are phenomenally aware of the full scene before us (at least at the centre of our attention). It seems to me it would take a potentially infinite amount of descriptions to describe all the properties and relations in front of you at any one point. Wouldn't it simply be more efficient to just register each 'visual pixel' with a cell so that more *general* descriptions can be formed further up the visual process (e.g. that green patch I am experiencing is called "grass" in English)? If this were true then it would seem that the propositionalist would also predict visual arrays at least in very early visual processing, and perhaps this will be most efficiently organised topographically at first. This is because the retinotopic data that is projected back onto the cortex will benefit from the same efficiency constraints. And indeed Pylyshyn (pg. 394) does not deny that there are topographic areas in the early visual cortex involved in vision. What he does deny is that these are necessarily depictive and are used in visual imagery. I have dealt with the former complaint already. The latter is discussed below, and is motivated by some studies that apparently show that these areas are not activated during visual imagery. So we now need to analyse whether early visual cortex is used in imagery and whether this relies on functional depictions in topographically organised areas. This is where we first start to analyse whether the visual buffer is also used offline to support visual imagery.

3.3 – Are topographic areas used in imagery?

As discussed in Chapter 1, a crucial claim of the Kosslyn Model is that the visual buffer is used in visual imagery as well as in vision. To counter this Pylyshyn often quotes empirical evidence that topographic areas in the visual cortex are in fact not used during imagery. For example, he specifically refers [pg. 394] to various studies by Mellet et al (1996, 1998), whose work will become relevant immediately below. Kosslyn replies to this critique in an extremely detailed fashion, by conducting a meta-analysis of 59 studies on mental imagery. This happens to include many of the studies that Pylyshyn also quotes. As part of the meta-analysis the studies were initially vetted to meet certain quality requirements and all involved only visual imagery, as opposed to imagery in other sense modalities. Each of the studies was then analysed and tested against the predictions of various theories. I wont go into the details of that analysis but merely report the relevant results to the discussion here (but see Kosslyn pgs.117-130 & Appendix).

An important initial criterion that Kosslyn sets up is that he is looking for activity in Brain Areas 17 and 18 to indicate activation of the visual buffer. This is even though his initial definition of the visual buffer is slightly wider including *any* topographically organised area in the *whole* visual cortex, which includes Area 19 also[53]. However he quotes Sereno et al. (1995) and their fMRI study to support the idea: 'that both these areas [17 & 18] are known to be topographically organised in humans.' (pg.117). Hence activation during imagery in either of these areas would indicate visual buffer activity, and conversely lack of activity together with reports of visual imagery would be a problem for his theory.

The main distinction which Kosslyn suggests needs to be made to explain the apparently problematic empirical data, is one between: imagery that involves considering high-resolution details of objects or shapes (depictive imagery); and those involving only spatial relations, such as comparing the relative sizes of objects, or where they are located in space (spatial imagery). An illustration of the potential difference between these is given below:

---

[53] Area 17 is known as V1 and the primary visual cortex; Area 18 as V2 and the secondary visual cortex, and Area 19 as the associative visual cortex and is thought to be used in processing long-term visual memories.

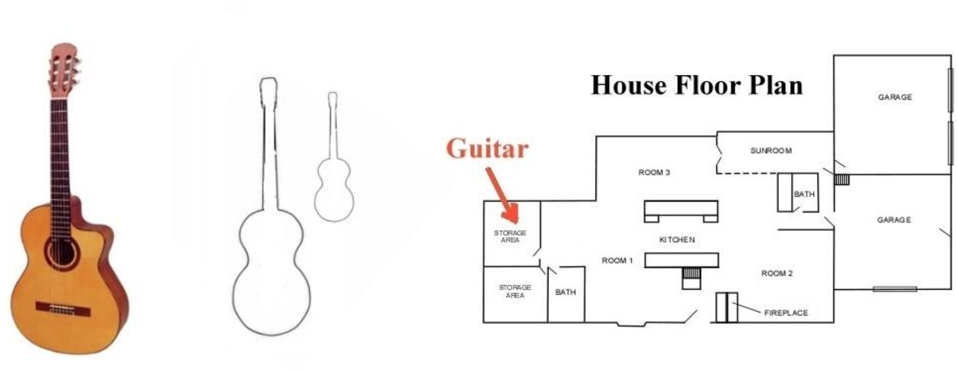


Figure 3.4 – Illustration of different styles of imagery: Left - detailed depictive image of a guitar; Middle - spatially comparing the relative sizes of two guitars; Right - imaging where a guitar might be spatially located in your house

As illustrated above, depictive imagery (left) involves the production and analysis of a high-resolution image of a specifically shaped object in order to gather some detailed information about it. This may also involve imaging the various sub-shapes and colours within the object in detail too (e.g. the strings, the frets, the tuning keys). Perhaps this could be used in a task that requires one to work out roughly how many frets a standard guitar has on its neck. Kosslyn interprets this as needing a detailed imagistic depiction of the object and this is what requires the activation of the topographic areas in the visual cortex. As Kosslyn puts it:

> '…the act of searching for a detail in a mental image results in actually placing the detail in the image, which apparently leads to activation in the topographically organised visual cortex. Alternatively, the image may need to be reorganised [i.e. scanned or rotated] to "see" a specific part in detail; this process would also require operating on a high-resolution depictive representation…' Kosslyn (pg.127)

Hence the requirement to imagine an object's shape and colour features *in detail*, in any imagery study would predict that the relevant areas of the visual buffer should be activated. Once this distinction was made, the meta-analysis found that cases of depictive imagery matched up with the studies that showed activation in the visual buffer.

Conversely, certain other studies that required only spatial imagery didn't show activation in the visual buffer. This is explained because, according to Kosslyn, the parietal cortex is responsible for processing this kind of information in its own topographically organised areas, and hence it does not necessarily need to utilise the visual buffer (we will look at spatial imagery in more detail in the next chapter). Kosslyn also points out that spatial imagery can involve quite high resolution and detailed inspection of spatial relations, and hence it is not the generation of high-resolution imagery per se that distinguishes these two kinds of imagery. It is more precisely whether it is a detailed image of shape and colour that is required, versus a detailed consideration of certain spatial features.

So given the distinctions introduced above we can return once more to deal with Pylyshyn's specific objections to the Kosslyn Model in this respect. It seems the some of the studies quoted by Pylyshyn include studies involving *only* spatial imagery and hence one wouldn't necessarily expect the visual buffer to be active anyway. Perhaps then the difference in opinion here is best explained by a refining of the Depictivist's predictions, based on the evidence, that the visual cortex will only be active during *depictive imagery*. To give a specific example, Mellet et al (2004) also conducted a meta-analysis of imagery experiments conducted by their lab in previous years based on PET scans. Recall that I have just mentioned that Pylyshyn appeals to many of the Mellet studies as counter examples to the Kosslyn Model. The Mellet meta-analysis returned results consistent with Kosslyn's predictions in terms of object imagery (their term for depictive imagery), and summarise this concisely in this exert from their abstract: 'object imagery activated the early visual cortex, whereas spatial imagery induced a deactivation of the early visual cortex.' Hence showing a specific way that the distinction between different kinds of imagery can defend against this line of critique from Pylyshyn.

What Pylyshyn now needs to continue with this avenue of attack, is to refine his critique so as to identify an imagery task that obviously specifically requires detailed depictive imagery, but where the visual buffer is still inactive. If this combination cannot be explained in any other way, then this would count as a new counter-example to the Kosslyn Model. However, for now, it seems that Kosslyn's recent approach is specifically designed so that it can deal with the current evidence and hence remains a live option pending future studies.

It might be noted that the guitar size comparison illustrated in Figure 3.4 still seems to involve imagining the relative shapes of the guitars, and if this is so, it might introduce a certain amount of a depictive element to this allegedly purely spatial imagery task. This anomaly can perhaps be explained in one of two ways. Firstly, it might be that this is just a feature of the way I have illustrated it here and it does not reflect what actually goes on in a real spatial imagery process. Namely, the guitar shape is not actually visualized in size comparisons and the above diagram just needs to depict shape to illustrate the point in a text-based explanation. Or secondly, the details of any shapes that are imagined in this situation are so minimalistic that the activation they require from the visual buffer is fairly undetectable above a background threshold. To that end, it would be interesting to read about an experiment designed to see if activation of the visual buffer increases with the amount of depictive detail required for a certain imagery task, and whether this effect disappears below a certain threshold.

Clearly there is still much refining work to be done in this area to predict which kinds of imagery task use which areas of the brain and how strongly they are activated. There is a real potential for even more fine grained distinctions to be made with which areas of the brain seem to process these even within the visual buffer (e.g. imagining faces vs.

imagining scenes; and see Chapter 7 for an example of this). But for now it seems the depictive vs. spatial imagery distinction predicts more accurately when the visual buffer is used, and why sometimes it is not used, in a way that is essentially compatible with most of the empirical data. We await the critic's reply on this matter and perhaps for now we can justifiably move on to the next issue.

Given the above, we can now *briefly* investigate the possibility that just because the visual cortex may be used in imagery *at all* doesn't mean that it relies on the topographic areas in the visual buffer to function. To see this more refined point we need to look at a study by Slotnick, Thompson and Kosslyn (2005) that actually shows these brain areas responding topographically in perception and equivalent kinds of imagery. This study was also designed to rule out various other criticisms from Pylyshyn as indicated in the quote below. In this study subjects were asked to observe and imagine a rotating chequered bow tie, while fMRI scans were conducted to detect brain activity. They were also asked to just pay attention to different parts of the display in order to detect activity associated with merely attentional demands. The results matched the predictions made by the Depictivist approach, both in terms of location and in mode of presentation: that is, not only were the topographic areas of the visual cortex used but the way they were used was similar in perception and imagery:

> 'The fact that the phase of the activity (representing stimulus position) in the imagery retinotopic maps was similar to that in the perception retinotopic maps — which in both cases directly reflected geometric properties of the stimulus — is in accord with Pylyshyn's criterion for evidence of depictive representation: 'that the way their topographical organization is involved reflects the spatial properties of the image' (Pylyshyn, 2002). In addition, our findings from both the main experiment and follow-up experiment counter Pylyshyn's claim that visual mental imagery is accomplished primarily through attentional mechanisms.' Slotnick, Thompson and Kosslyn (2005: 1582)

Hence this represents the latest in evidence that responds to scepticism from Pylyshyn and provides further empirical evidence in favour of Depictivism. Not only are these topographic areas used in a way that is different than attentional demands, but they are also activated in a similar way in vision and visual imagery. That is, seeing a rotating chequered bow tie utilises the visual buffer in a similar way to the geometric demands of a similar task in imagery. Of course, given that the strength of the stimulus and the vividness of the experiences are typically much lower in mental imagery, it predictably has a much weaker effect, than in the perceptual case. Again we await the response to this evidence from the propositionalist camp.

Although the above shows that the same topographically organised areas are used in similar ways in both vision and imagery, it doesn't necessarily show that it is the same mechanism that is used in both. It is possible that imagery just uses areas in the same vicinity as the perceptual process, but in fact they just happen to be located near enough to each other, and it just so happens that distinguishing these two separate streams falls below the resolution of the best modern brain scanners. However, as mentioned in Chapter 1, data from Kosslyn et al (1997 & 2001) show that visual imagery interferes with visual processing, thus suggesting they are competing for the same resources. This is opposed to auditory imagery that does not interfere with visual processes, which also indicates that it is not just a phenomenon that derives from a competition for our limited attention. Hence there seems to be fairly compelling evidence to show that not only are some of the same areas used, but that they also compete for resources available in the same mechanism. This is a correlation that is strongly predicted of, and elegantly explained by, the imagination-as-simulation hypothesis, and indicates that the visual buffer is therefore a good candidate for an online main visual mechanism being used offline to support visual imagery.

It is worth noting here, that this identification of a main mechanism does not imply there is complete overlap in all respects, since obviously there are differences in other areas of these two processes e.g. where they process their (mock) inputs and (mock) outputs. And analysing the amount of overlap between vision and visual imagery links with another major avenue of critique available to the Depictivist camp that I would like to address in the next chapter. This comes from purported evidence that people who have damage to the visual buffer can still form visual images. I think the best way of replying to this is to first give Kosslyn's full positive account of the visual imagery system and how this can be affected in different ways. This will help explain how they deal with the apparently contradictory dissociation data observed from patients with brain damage. Again the Kosslyn Model described in the next chapter is specifically designed to be able to account for this data. But before going onto that there are a few final general points to make.

3.4 – Remaining resistance to depictive representational approaches

Given the above, we might want to assess how strong the resistance still is to at least partially depictive approaches in the broader imagery research community. Kosslyn (pg. 179-80) addresses this point in the final chapter of his book and reports back on a survey from 150 researchers in this field compiled in 2003. Firstly he reports that only about 7% actually favour a purely propositionalist account, with the majority accepting some depictive representations are necessary. Secondly he suggests that those who reported not having any visual imagery on a questionnaire, were more likely to be propositionalists at the onset of the imagery debate in the 80's. Indicating that one's personal phenomenology might initially guide how one interpreted the debate. Subsequently, many of those who report not experiencing phenomenal imagery, have changed their minds to a more mixed theory based on the accumulating empirical evidence in favour of Depictivism; with only a few remaining outright sceptics who deny depictive representations are used at all. The point of mentioning this is just to flag up the possibility that variations in phenomenology might strongly bias the initial position you take in this field. Hence we must be wary of phenomenological reports in analysing this debate and take care to separate the objective facts from subjectively biased reports[54]. But this feature may explain why there was such an initial difference in the claims of these two approaches that has slowly narrowed as the evidence accumulates.

Countering that thought, it might perhaps also be suggested that the apparent gradual acceptance of depictive representations described above, could mainly be due to Kosslyn weakening his position so far that it is now largely compatible with Propositionalism (e.g. the move to a hybrid approach). But this can equally well be seen as a continued modification of his theory in reply to criticism and a refining of the empirical evidence. Which is a perhaps a natural and healthy scientific methodological process; that is as long as this doesn't begin to look a bit ad hoc, which I don't think it does. The evolution of his ideas towards a hybrid approach, perhaps also means that the benefits of the predictive and explanatory power of his approach could be available, with certain modifications, to those that would usually emphasise the propositional element. It certainly seems that there is less to divide the positions than in previous versions. As mentioned earlier, my aim is not to particularly to decide between the theories, since the simulationist camp can live with both. My aim is more to get a detailed explanation of what might be happening in sub-personal mechanisms serving vision and visual imagery, and I think Kosslyn's hybrid theory is the leader in the field in that respect.

To highlight the benefits of a Depictivist approach further, perhaps we can look briefly at the positive proposals from the propositionalist camp. As discussed, they are unable to

[54] And perhaps this may also affect your initial position in terms of the selectivist vs. generativist debate as introduced in Chapter 1 and to be discussed further in Chapter 5.

make testable predictions on the scanning and rotation tasks, and can only explain the data post hoc by invoking new-list-properties. Alternatively they can try to explain away the results by appealing to external factors, such as tacit knowledge or cognitive penetrability arguments. Kosslyn adds to these remarks as follows:

> 'We were tempted to try to derive predictions about which specific areas [in the brain] would be activated during imagery if a propositionalist theory is correct. This proved impossible, in large part because no detailed version of propositional theories of imagery have been developed… this class of theories is distinguished in large part by what it rejects rather than what it posits.' Kosslyn et al (2006: 121)

For example, much of Pylyshyn's book is devoted to a negative critique of Depictivism (Chapter.7) and developing arguments against the functionality of phenomenal mental imagery (Chapter.8), and relatively little to explaining how Propositionalism can positively explain the data. The utility of this approach at all levels, is supposed to be obvious from its ability to deal with higher thought processes based on 'language of thought' arguments. However to some, this extrapolation is not very clearly achievable at all and it tends to set the propositionalist against those with a more empiricist or dual coding approach (c.f. Prinz 2002, Paivio 2007). For example, this scepticism is described well by Barsalou (1999: section 1.2.2) who suggests there is at best only indirect evidence that propositional-style amodal symbolic representations exist. They are a theoretical entities invoked to describe our ability to understand and deploy concepts: a mode of representation that is then assumed to occur ubiquitously in all forms of thought. However it is possible that modally specific representations, like visual imagery, can do a lot of the work usually ascribed to propositional representations. So the debate very much boils down to whether you think that propositional 'language of thought' style explanations can explain all the data or whether you think there is at least some depictive element involved.

In my opinion I find it hard to see how that a finite set of propositions could describe a phenomenally rich visual scene and all its complex spatial relations adequately in the first place. For instance, wouldn't we need a proposition to describe the colour of all the discernable small areas of the visual field? This would be required so that we can at least answer the following series of questions about our phenomenal experience: what is the colour of *that* area? …and also the next area along? ..and the one next to that? …and so on. It seems to me that this is best done by describing each small area (or pixel) of the scene, which would also need to be represented in a corollary set of cells in an array. These cells can then be variously accessed to provide information about other, more coarser grain, descriptions of the objects and their relations within the scene: e.g. the red ball is ON the blue box. But a cell-by-cell propositional-style description of the visual scene just seems to me to be broadly the same as a functional hybrid depictive representation. And note that this is the same form of argument as that given above, about having a padded cell list of all possible directions you could scan a map. Hence I suspect there is some overlap between these two ways of approaching the subject, with the depictive representational approach perhaps now offering more explanatory power and empirical consistency besides. Therefore at this point I feel fairly justified in adopting a hybrid depictive approach for the remainder of the thesis. In the next chapter we will look at more detailed aspects of the Kosslyn Model and how it can explain apparent dissociations between functioning vision and dysfunctional visual imagery (and vice versa).

4 - STIm and the Full Kosslyn Model

> 'Imagery is like a juggling act in which only a small number of balls can be kept aloft at once.' Kosslyn et al. (2006: 42)

> 'Mental simulations are imagined scenarios that mimic what one would expect to happen in the corresponding actual situation, and depictive representations play a key role in such reasoning because they make explicit and accessible aspects of shape and spatial relations that otherwise need not be evident. For example, consider what occurs when you have a pile of luggage and must pack it efficiently into a car's trunk. You can visualize the bags in various positions and locations, mentally moving them around until you "see" a good configuration – all before you even begin to haul a single bag off the ground.' Kosslyn et al. (2006: 71)

In this chapter we will look in detail at Kosslyn's full account of the visual imagery system, which is described in Chapter 5 of Kosslyn et al (2006). As discussed in the last chapter, the visual buffer is thought to be a crucial part of the Kosslyn Model, but there are also other important mechanisms that work in conjunction with it and these will be introduced and analysed here. So while in the last chapter we mainly looked at what a hybrid depictive representation actually is and at evidence that topographic arrays in the visual cortex were used in visual imagery, in this chapter we will now look at how the visual buffer connects with other mechanisms in the brain. From this we can then see if this seems like a model that can be interpreted as an offline simulation process and whether we can identify other mechanisms that we would expect to find in this kind of a system. We will also look at, and reject, some further putative counter examples to the necessary involvement of the visual buffer in depictive visual imagery.

So specifically, in this chapter I will try to do the following three main things. **Firstly**, it will give a fairly full account of the latest version of the Kosslyn Model and explain how the visual buffer is proposed to link up with other important mechanisms. **Secondly**, I will investigate some strong putative counter examples to this model that seem to suggest that imagery can persist even without an intact visual buffer. If this is correct it represents knock down argument against the Kosslyn Model. However I will argue that once the details of model are explained, the force of this evidence is more apparent that real and hence can be dealt with adequately. **Finally,** we will look at how the visual buffer interacts with other mechanisms in detail and how this can be seen as an offline simulation process. In this chapter I will look mainly at how the inputs to the visual buffer are gathered together and delimited, and I will postpone an analysis of the output processes for later chapters. Analysing the input side will involve, amongst other things, looking at possible ways that the contents of the visual buffer can be filled-in *internally* by other areas of the brain. This is as opposed to being filled-in by *external* stimulus from light detected at the eye. By the end of this chapter we should have a fairly detailed and robust theory about the underlying processes supporting phenomenal imagination and hence STIm. This will be used later to underpin a more full account of how we can phenomenally imagine other times and places. In this chapter we will also specifically develop a theory of how we can imagine a certain spatial layout, and the *temporal* aspect of Spatio-temporal Imagination will be developed in subsequent chapters. I will start now by just simply introducing and explaining the main mechanisms in the full Kosslyn Model.

4.1 – the full Kosslyn Model - a theory of the visual imagery system

One of the most crucial features of the Kosslyn Model is the bifurcation of the visual stream into the *dorsal* and *ventral* streams. Kosslyn suggests that the **dorsal stream**, which runs upwards along the back of the brain (dorsally) from the visual cortex into the parietal cortex, is responsible mainly for spatial properties processing (SPP). Hence this stream mainly deals with the spatial aspects of vision and certain features of this have been touched on in the last chapter, where we discussed the nature of spatial imagery (e.g. locating a guitar in your house relative to where you are and comparing the sizes of two objects). Relatedly this stream is also sometimes referred to as the “where” stream, because this involves processing information about where objects are located relative to the perceiver. It is also sometimes referred to as the “how” stream, as it is involved in coordinating motor actions. This stream is thought to involve the use of an *egocentric coordinate* system that processes information in a coordinate system that is centred on the self, which we will discuss in more detail below. This stream is also involved in manipulating map-like representations of where you and other objects are located. This allows us to keep track of where we are in our environment and where you may need to go to do certain things. We will look at some examples of how this may work in Section 4.3.

The **ventral stream** on the other hand, which runs directly forwards in the brain (ventrally) into the temporal cortex, is responsible for recognising objects and object properties processing (OPP). This is also sometimes referred to as the “what” stream, because it allows you to recognise and determine what certain objects are. The *object properties* it processes are things like colour and shape, but not necessarily where it is spatially located relative to the observer. The separation of these streams therefore allows dissociation between naming and recognising objects in the OPP ventral stream and locating and manipulating objects in the SPP dorsal stream. This bifurcation of streams is fairly well accepted distinction in mainstream psychology and hence will not be questioned here[55].

---

[55] Although Kosslyn recognises that Milner and Goodale (1995) have suggested the parietal cortex is mainly used to help guide movement (how), they prefer to adhere to the original notion put forward by Ungerleider & Mishkin (1982) that is responsible for broader spatial (where) processing. However they suggest the two are not mutually exclusive: one needs to know where something is in order to understand how to use it. Hence the areas controlling these functions may overlap in the parietal cortex. More on this later and recall the approximate locations of these streams and the main areas of cortex are depicted in the diagram in Appendix B.

The main support for this bifurcation comes from observations of what happens when these streams are not working properly. For instance, damage to the OPP stream, classically due to damage to the temporal cortex, leads to various forms of Agnosia. This is where, for example, people cannot name objects, or recognise people or recognise text. Conversely, damage to the SPP stream, usually because of damage to the parietal cortex, leads to various form of Neglect. This is where half of the external world is ignored or neglected, or is not represented in the first place. Damage to the parietal cortex also leads to problems with motor coordination (e.g. reaching and grabbing actions) and discriminations involving sizes and locations. These distinctions will be important later and represents the main division of labour within the Kosslyn Model.

In addition to these two streams and the visual buffer, Kosslyn invokes three other mechanisms in the full Kosslyn Model. The first is **Attention Shifting**, which can change the focus of attention in vision and visual imagery. This has the effect of moving an *attention window* within the visual buffer so that one can concentrate on different things within the focus of that window. The second mechanism is that of **Associative Memories,** which takes inputs from both streams and is used to identify and re-identify objects. Long-term memory is linked with several areas, including the anterior temporal lobe, Wernicke's area and the "association cortex" (Area 19 of the visual cortex), which are all in the temporal lobe. Conversely, short-term memory (working memory), which only lasts for a few seconds, is linked to the prefrontal cortex, which is in the frontal lobe. We will be discussing mainly the use of long-term memory in this context, and will investigate its use as a mechanism that helps guide what we visualize relative to a certain task. For instance, if we want to visualize a Rubiks Cube we may need to retrieve its parameters from long term memory to delimit the shape and colour of our eventual visual image of it. Again this illustrates the close link between memory and imagination in this model, since we can not only *remember* the shape of a Rubiks Cube, but we can also creatively *imagine* what it might look like if it was bright pink all over. Hence this model fairly seamlessly integrates memory and imagination components in one visual imagery system. And this is part of the reason why I didn't worry too much about untangling these two features in Chapter 1.

If an object is not immediately identifiable, and if identification is important, then a third mechanism called the **Information Shunting** mechanism sends relevant information to the other systems to shift attention towards certain features of that object. This helps *prime* lower areas for key things to search for. Hence it interacts with the Attention Shifting mechanism to conduct a relatively slower process of object identification, which is referred to as a *search priming process*. This may be used mostly when an object is seen from an unusual (non-canonical) angle: for example, a fence line seen from directly above or a guitar seen from the end of the neck. The Kosslyn Model is summarised in the diagram below:

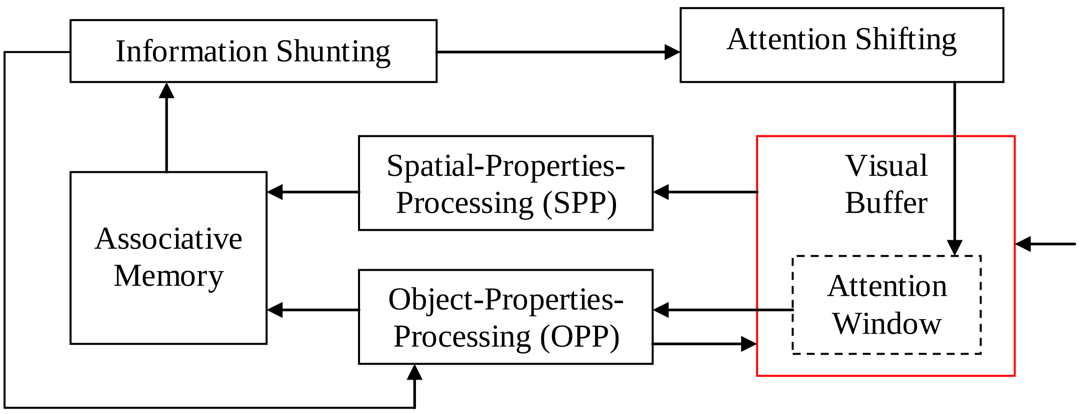


Figure 4.1 – the 'Kosslyn Model' of the visual system:
Based on Figure 5.1 in Kosslyn et al (2006: 136)

The Kosslyn Model predicts that some of these mechanisms and associated brain areas will also be active where appropriate in a visual imagery task. It particularly predicts that invoking a visual image is a bit like the *search priming process* described above, but where the priming surrounding the visual buffer becomes so strong that an actual image is produced. This process of filling-in the visual buffer will be discussed in detail in what follows, as it may suggest a way of describing how the visual buffer gets fed mock inputs in simulationist terms. Obviously this system is designed to fit with all the known empirical data and specific examples of this will also be discussed below and in Section 4.3.

In terms of the visual image itself, there are four *processes* that are thought to be needed to explain our ability to utilise it under this model. These function to do the following things:

1) Image Generation – this uses propositional information stored in memory areas to reconstruct or generate visual images that make the shapes and colours of what is imagined explicit. If the task requires considering detailed shape, then this could involve information from associative memory to be sent back along the OPP stream, to the visual buffer to generate the image of these object properties e.g. the colour and shape the object takes. If however it is just spatial imagery that is required, then it may be that only parts of the SPP stream are utilised. Since these areas specifically retain spatial information in topographical arrays they do not need to access the visual buffer to make this information explicit. However data from the SPP may be used to help control the spatial layout of a depictive image (as discussed in the last chapter and explored further below).

2) Image Inspection – this shifts the attention window within the image but also can lead to image generation in new areas that need to be attended to. Hence it sometimes works in conjunction with image generation, as you realise you need to attend to a new area, or to attend to something in more detail. At this point the required part of the image then needs to be generated accordingly.

3) Image Maintenance – normally in perception new information from subsequent saccades of the eyes over writes previous information in the visual buffer. Hence representations are not designed to last in this medium for long and this is why a depictive image begins to fade quickly. However by concentrating on maintaining the image, information on what is being visualized can be re-sent to the visual buffer and imagery maintained. This takes some effort and explains the phenomenology of imagery in this respect.

4) Image Transformation – this allows the image to be manipulated such as in rotation and scanning tasks. Kosslyn predicts there are two varieties of this process. One where you imagine how an object would change due to you manipulating it physically (e.g. rotating it with your hand), and the other where external forces might change how an object looks (e.g. gravity making it fall). They predict that only in the former case will motor areas be activated as it simulates action commands. Both processes involve changing the mapping function between the OPP stream and the visual buffer and controlling how it is manipulated and hence changing how it would now appear to you.

Table 4.1 – Mechanisms needed to examine and manipulate a visual image under the Kosslyn Model

We will return to discuss most of these imagery specific mechanisms in more detail below, although we will not have time to look at the image transformation mechanism in detail[56]. Particularly, in this chapter we will look at the steps involved in the input side of the Image Generation process, especially in Section 4.3. For now we can observe that the Kosslyn Model involves the interaction of many different mechanisms in order for it to work correctly and because of this there are potentially a lot of different things that these can go wrong independently of each other. This flexibility in the ways that this system can break down will be crucial to explaining certain imagery deficits in Section 4.2. It should be noted that the complexity of this model is shaped by the available empirical data and is specifically designed to allow for this flexibility in explanation.

---

[56] A detailed section on this was removed from the thesis to keep to the main focus of specifically analysing STIm. Suffice to say that there are some interesting ways of explaining how the motor areas may interact with visual imagery to aid image transformation in certain cases: e.g. rotating an object in the horizontal plane, is a bit like imagining walking around it. And this potentially offers a way of explaining how these capacities arose in the first place and is linked to the analysis of our ability to imagine other places in the analysis below.

Having said that, a full analysis of the data motivating this model would go beyond what I can do here, but to illustrate the most relevant points we can look at two crucial aspects. **Firstly,** we will look at evidence for *overlap* between the mechanisms used in vision and visual imagery under two seemingly different tasks, which are however predicted to use similar areas under the Kosslyn Model. **Secondly,** we can look at evidence from their *dissociation* where vision is lacking but visual imagery seems to persist, even though the visual buffer is severely damaged. Obviously identifying a single case of persisting visual imagery without a visual buffer would be a knock down counter example to the Kosslyn Model. Hence formulating a reply to it represents the crucial defence of this approach against what I think is the strongest challenge to it so far. This will be the final putative counter example I will deal with, before accepting the Kosslyn Model wholesale and utilising it more positively in the second half of the chapter. We can now explore how the Kosslyn Model is positively motivated and how it deals with some seemingly threatening counter examples. In doing so we will also get a better idea of how the Kosslyn Model is thought to support visual imagery.

4.2 – Evidence of overlap and dissociation

4.2.1 - Overlapping evidence

In the last chapter we looked at evidence from some experiments suggesting overlapping use of *specifically* the visual buffer in vision and visual imagery. In this section we will look at a more widespread overlapping use of the Kosslyn Model as a whole system. To show how robust their approach is various experiments have been designed by the Kosslyn Lab to test their model[57]. These involved conducting experiments on vision and visualization tasks that on the surface of it appear very dissimilar, but are predicted to show

[57] The experiment described below is from Kosslyn, Thompson and Alpert (1997), but Kosslyn (pg.151-157) refers to many other papers to back his point up, and his lab has produced many of these papers.

major overlap if their model is correct. For instance, as introduced above, the Kosslyn Model predicts that in **vision**, when an object is not immediately recognised, areas responsible for 'Information Shunting' will send signals to prime other areas in the OPP & SPP streams, and prepare them to look for certain features in the visual buffer. Hence showing subjects non-canonical pictures of otherwise familiar object (e.g. a guitar from the end of the neck) can potentially stimulate this *priming* process. It was therefore predicted that all the visual mechanisms described above would be activated in this task.

In **imagery** on the other hand, it is suggested a similar priming process occurs except in this case it is: 'strong enough that the representation is evoked in the early visual cortex, reconstructing the shape of the object in the visual buffer…' (Kosslyn: pg.155). The main difference here is that you are not trying to categorise a perceived object, but generate an image of an object that is detailed enough to complete a task (e.g. the Rubiks Cube Problem). To test this second imagery case, they asked people to imagine the upper case version of a letter and asked them to judge whether a target X would fall where that letter would be on a 4x5 grid. See below for an illustration of this experiment:

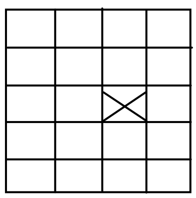
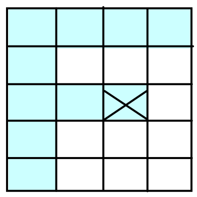
1) Task: Imagine an upper case letter 'f' and answer the following. Would the x in the diagram adjacent fall on it if it were drawn there?	2) 4x5 grid and cross only: 	3) Grid with "imagined" upper – case letter superimposed:  Answer: yes it would.

Figure 4.2 – Illustration of imagery task described above:
from Kosslyn, Thompson and Alpert (1997)

Again it was predicted that all the main Kosslyn Model mechanisms and processes would be activated. So on the surface of it these seem like very different visual and visualization tasks: i.e. non-canonical object recognition vs. imaging a letter. Both seemingly involving very different mental capacities, but given the Kosslyn Model it was predicted that these processes would significantly overlap in brain areas activated.

The results from the experiment broadly confirmed these predictions with some useful qualifications. On the whole there were 21 areas activated and 14 of these were activated in common, and all were in the expected areas of the brain. In addition 5 areas were only activated in the imagery task and 2 areas activated solely in the visual task. But although slightly different areas might have been used, these were in fact classed as different areas of the same mechanism or stream. Suggesting the same processes were used but in slightly different ways, which seems like a natural way to interpret the situation.

In detail then, there were two areas associated with the SPP stream that were activated in both tasks, but there were also two areas associated with this stream that were activated only in the imagery task (50% disparity). Hence although the SPP stream was activated in both cases the actual way it was activated was slightly different. Additionally the OPP stream and the Associative Memory mechanism had 33% disparity, the Information Shunting had 25%, and most importantly the visual buffer shared all areas (no disparity). The latter result indicating that the visual buffer in *general* was used in both cases, although I assume there was a difference in strength of activation with the visual case being greater than the visualization case (see below). The predictive power of this test not only supports the general hypothesis, but the details of the areas activated also supports the individuation of the mechanisms. I am unaware of any specific objections to this evidence, other than general challenges to the Kosslyn Model, which are discussed next.

4.2.2 - Dissociation evidence

The main challenge from the brain damage literature in this area comes from Bartolomeo (2002), who reviews a number of these cases and suggests there are some examples that put pressure on the Kosslyn Model. His first claim (Section.1) is quite weak, and suggests that in isolated imagery deficits, with no associated visual deficit, it is the temporal cortex that is always damaged in some way. However this from of dissociation is consistent with a problem with *image generation* in the Kosslyn Model rather than necessarily requiring a separate 'imagery buffer' in the temporal cortex as Bartolomeo claims. In other words this can be explained because the correct stimulus for generating an image from associative memory is not available and this could be due to damage to the temporal or the visual cortex, or both; the results will be the same as both are needed for depictive imagery. Hence an imagery deficit in a subject with an intact visual buffer is explained by an upstream problem with image generation consistent with temporal cortex damage that interferes with the OPP stream. These results therefore do not seriously challenge the Kosslyn Model.

More challenging are reports from people with cortical blindness[58] and extensive damage to the visual buffer areas, but who still retain some form of visual imagery ability (Bartolomeo 2002 Section 2.1). In response to this Kosslyn (pg.165) refers to detailed fMRI and PET studies conducted by his lab[59], which indicate that although the visual buffer is active on both cases, generally less of the it is used in imagery than in perception (by about 25%). And further, that imagery activates the visual buffer much less strongly than in perception[60]. Leading Kosslyn et al (2004: 234) to suggest that: 'Clearly, sensory

---

[58] This is where the eyes are responsive and functioning and blindness is due to more upstream brain damage, usually to the primary visual cortex (V1: Area 17).
[59] For PET scan results see Kosslyn, Thompson & Alpert (1997); for fMRI – Ganis et al (2004 -especially pgs. 234-5).
[60] Unfortunately I have been unable to find a percentage figure to put to this general claim.

input drives [the visual cortex] more strongly than input from information stored in memory'. This may be because in perception more processing is needed from the raw sensory information (e.g. edge detection), and the stimulus is much stronger when generated externally rather than internally.

This feature seems to match up with first person reports that *generally* vision is more vivid than imagery and hence is not really a surprising result. In contrast, with imagery the main features of basic objects may already be known and just the areas responsible for more general upstream processing may be required. For example, it may be that edge detection is not required to generate images from memory as this has already been processed in the original perceptual case. Hence it is *possible* for early processing areas in the visual buffer to be damaged and cause cortical blindness, while the areas required for imagery and more upstream visual processing to remain intact. It is tempting to suggest that Areas 17 and 18 are used in vision, but only the upstream Area 18 is used in visual imagery, as this would fit nicely with this kind of explanation[61]. However this is not what Kosslyn claims and may be far too simplistic given the complex anterograde (forward) and retrograde (backwards) connections in the brain. Instead we may just have to say that as long as there are some parts of the visual buffer intact then it seems that visual imagery may be still be possible even though vision itself is prevented.

To illustrate this point further lets look at a specific example of this challenge to see how well the Kosslyn Model deals with it. Bartolomeo (2002: 363) refers to a study by Goldenberg (1995) of a subject (H.S.) with Anton's Syndrome. This involved cortical blindness due to *almost* complete destruction of the *primary* visual cortex (V1: Area17), combined with vivid hallucinations such that the subject sometimes felt she could really

[61] Reminder: Area 17 is known as V1 and the primary visual cortex; Area 18 as V2 and the secondary visual cortex, and Area 19 as the associative visual cortex and thought to be used in long term visual memories.

'see' things. Complete denial of ones syndrome is referred to as Anosognosia and Anton's Syndrome is characterised by claims of perception, followed by a re-arranging of the facts to explain cases where they find they get it wrong. In other words, they concoct stories to explain why their "vision" let them down rather then accept they are blind and experiencing hallucinations. The pertinent facts to our discussion here is that H.S. still potentially had some areas of the visual buffer intact. Specifically H.S. had some unilateral (one side only) damage to Areas 18 and 19; and even Area 17 had isolated areas that seemed intact. This was to such an extent that H.S. eventually regained a small part of her vision. Which gives an indication of how far from complete the damage to her primary visual cortex was. Goldenberg summarises the situation like this (and see also pg. 1381 for a similar exposition):

> 'MRI showed an almost complete destruction of primary visual cortex with sparing of only small remainders of cortex at the occipital tip of the left upper calcarine lip. In the literature there are a few cases of denial of blindness with similarly severe damage to primary visual cortex but none with unequivocal evidence of complete destruction of primary visual cortex. We conclude that severe damage to primary visual cortex is compatible with visual imagery but that there is a possibility that islands of visual cortex must be spared to permit the generation of mental visual images.' Goldenberg (1995: abstract)

So it seems this was the most extreme example of Anton's Syndrome in the literature at that time and yet this makes no real impact on the predictions of the Kosslyn Model. I would suggest that this is due to a lack of engagement with the details of the Kosslyn Model that predict more upstream areas of the visual buffer (e.g. Area 18) *can* be utilised for visual imagery. It will be interesting, as briefly alluded to in the last chapter, to see if the Kosslyn Model can be refined so that certain areas of the visual buffer can be associated with different kinds of imagery and perception, depending on the details of the task: e.g. does it require edge detection or just general colour and shape generation. That is of course presuming that these can in principle be isolated and detected at such levels of detail. If this is possible it would allow further empirical work to attempt to detect more refined dissociations of what can be seen and imagined based on differential damage *within* the visual buffer itself. So this line of attack seems ineffective once the details of the Kosslyn Model are engaged with, but could lead to motivating finer grained refinements of the model in the future.

However there is one last strong challenge from Bartolomeo (2002: 365) that I think gives another indication of the intricate nature of the current debate. Here he describes a subject Madame D who has different forms of Visual Agnosia (can't name objects, faces, letters etc), but very clear, effortless and vivid imagery in the same domains. So much so that she complains of unnecessary tests on her imagery capabilities when it is clear she has a problem with her vision. In other words she is well aware of her condition and hence not also Anosognosic. Madame D has damage to Areas 18 and 19 bilaterally that extend to the temporal cortex on the left side, but importantly Area 17 is preserved.

Bartolomeo suggests that it is only possible for Kosslyn to explain this dissociation by very specifically isolating damage between two processes that he suggests should be physically close to each other in the brain. These are: on the one hand, the projections *from* the visual buffer to the temporal cortex (forward or anterograde projections); and on the other hand, those involved in the *reverse* process in imagery (reverse or retrograde projections). Bartolomeo suggests that damage to one stream and not the other is quite unlikely as the location of the 'cerebral white matter' supporting these projections are so close together that they are more likely to be damaged together. In other words it seems *unlikely* that bottom-up perceptual processing would be so extensively damaged, while top-down imagistic processing would be entirely preserved. One would expect some deterioration of

Madame D's imagery capacities, but none is evident. This argument is possibly even more of a challenge than Bartolomeo thinks, because he assumes that the intact Area 17 explains why imagery is possible at all, and why he focuses on the links between the visual buffer and the OPP. Where as from the above we can see that actually bilateral damage to Areas 18 (also visual buffer) and Areas 19 (in the OPP stream) *might* make matters worse for the Kosslyn Model.

Kosslyn does not respond to this point directly, but I think that a response is possible. This is because although damage that is quite this specific is *unlikely*, it is not impossible. And given the example of overlapping areas above, we can see that differential damage and subsequent dissociation is allowed for in the model. Hence Kosslyn suggests: '…we would be surprised if imagery and perception could *not* be dissociated. As we have seen previously, they are not identical; rather, the two functions draw on most- but not all –of the same brain regions.' (pg.165). Hence the criticism discussed here may just be an example of an unlikely freak combination that shows this dissociation potential in its most extreme and refined form. It may also be that there are different paths available for image generation that aren't located so near the forward projections from the visual buffer to the temporal cortex. For example it may be that the backward paths he refers to are actually more essential for the non-canonical recognition process rather than for imagery, and image generation is a relatively more plastic and distributed pathway. So although the areas used for imagery and vision *generally* overlap, this does not mean they cannot be damaged differentially in rare cases.

What Bartolomeo and any other critic needs to do in order to disprove the Kosslyn Model, is to show that high-resolution depictive visual imagery can persist with *complete* destruction of the visual buffer; that is with complete damage to all topographically organised areas in the visual cortex. No such study exists I presume or it would have been discussed in the literature by now. So for now we can suggest that the Kosslyn Model can cope with the known empirical data due to the intricacy of its conception. Which of course is no accident since this is exactly what it was designed to do. Hence confirming its principled construction (i.e. its not ad hoc or post hoc) and its considerable predictive power, testability and robustness.

For someone like Bartolomeo who still thinks that visual imagery is more likely to be dependent on a separate 'imagery buffer' in the temporal cortex, there also remains the challenge to explain away the positive data referred to by Kosslyn. Which is that the visual buffer is always active in depictive imagery. If it is not essential, then why does it fire? And note that arguments suggesting it is epiphenomenal would be fairly weak, since as we have discussed in the last chapter, the brain is a power hungry organ and unnecessary firing of neurons would be far from epiphenomenal. This is because it would presumably incur a severe disadvantage to an organism in terms of wasted energy. So much so that it is unlikely that this excessive firing would have avoided de-selection over the generations. I would suggest that the associated firings in the visual buffer are much more likely to be functional, and most likely it is functional in a way that mimics its well-established online function of sub-serving visual style experiences. If it supports bottom-up *online* visual experience then most likely, all other things being equal, it is a good candidate to support top-down vision-like experiences when it is re-used for another purpose *offline* (i.e. visual imagery).

For me, it makes sense that as long as the visual buffer contains the right information to support vision (e.g. shapes, depths, colours), then this way of representing 3D objects could also be parsimoniously used to support visual imagery *in some fairly similar way*.

Why bother forming a whole new brain area to do this? Especially in an area like the temporal cortex that is set up to name objects and form memories, and not to process hybrid depictive data. It is worth noting here that I assume agnosic subjects do actually have a phenomenal experience and actually ‘see’ the shape of the objects, its just they cant recognise or name them. This is supported by the fact that they don’t complain of blindness, just an inability to recognise and name things. Hence while it may be classed as a general perceptual recognition problem, it may not be a problem with generating phenomenal experience in the first place. Hence it seems that the visual buffer can be further isolated as essential for vision-like phenomenal experience. These last points also give us an intuitive reason for favouring the Kosslyn Model: supporting phenomenal visual imagery just seems a priori like the right function for the visual buffer to be involved in, since it also seems crucial in supporting our online phenomenal visual experience; and further it seems that the temporal cortex just seems to be specialised in a different task and hence is a less likely candidate for underpinning visual imagery.

We can also suggest that the Kosslyn Model is so well entrenched and tuned in with the current empirical data, that barring a whole-sale rejection of many of the supporting experiment’s validity, then it is more likely to need slight modifications than suffer outright falsification as new empirical data emerges. This modification may take the form of: defining exactly which areas of the visual buffer are used in certain forms of imagery; which parts of the SSP and OPP streams are used when; and more detail in how the “spatial buffer” works and interacts with the visual buffer (see also below). So on the whole this looks like a theory that has the potential to fully explain the mechanisms involved in visual imagery in quite fine detail and is therefore, in my opinion, a sensible position to adopt at present. I certainly can’t really compete with Kosslyn’s 30yrs experience of study on this subject, but have hopefully done just enough to satisfy myself (and hopefully the reader) that this theory of mental imagery is very plausible and hence naturally becomes leader in its field. Therefore I will assume I am reasonably justified in adopting it for the remainder of the thesis. While at the same time accepting that this is a foundation that can be returned to and questioned at any time in the future. So from this point forward I feel I have defended the Kosslyn Model in enough detail to justify adopting it without qualification in what follows. At any rate the veracity of the model will be the assumption for the remainder of the thesis and hence will be used as a platform for further analysis.

4.2.3 – Dissociation evidence from Neglect and the SPP stream

As a way of summarising and generalising the form of the argument given immediately above, we can perhaps now look at how it might deal with dissociation evidence from Neglect due to damage to the SPP stream. The *general* form of the argument above, as I see it, is that there are at least two ways that vision and/or visual imagery can be variously affected. The **first** way is where an involved mechanism is damaged and hence cannot operate as it is supposed to. This could be due to damage to the visual buffer itself or a mechanism that usually provides relevant information to it. The **second** way is where the communication paths between two involved mechanisms are disrupted and information that is being passed between the two is prevented from doing so in some way. This would involve damage to the forward or backward projections between the two mechanisms as discussed above. In addition to this there are various permutations of each type of damage where only partial or one-way damage occurs. For instance, a mechanism might only be partially damaged so that it can do some operations but not others, or the connections between the two mechanisms are only damaged in one direction or only partially damaged in one or both ways. Finally we can envisage situations where there may be a complex and subtle combination of both types of damage. For example, where there is partial damage to the visual buffer and partial damage to its network of connections to other mechanisms. The sum total of these permutations means that there are many different ways that this system can go wrong and hence this re-enforces Kosslyn's claim that it would be surprising if we did not see dissociations between our visual capacities and our ability to generate and manipulate visual images.

We can now perhaps apply this general line of argument to specific cases of damage to the SPP stream, and hence potentially explain double dissociations in the field of *Visual Neglect.* This is where one side of the visual field isn't experienced and there are corresponding cases if neglect in imagery where one side of a visual image is not experienced. This latter form of imagery specific Neglect, is referred to as *Representational Neglect*[62]. As mentioned above the main disorder associated with problems with the SPP stream and damage to the parietal cortex is generally referred to as Neglect, because it sometimes leads to the affected individual neglecting a certain part of their full potential experiential range. Due to certain processing features of the brain it is most commonly the left hand side that is neglected, and I will only refer to this variety in what follows[63]. In extreme cases of Neglect individuals will fail to shave, or put make up on, the left hand side of their face. And when combined with Anosognosia, they may deny that their left hand or arm belongs to them, or that there is anything wrong with their limited visual processing. For example, they may try to throw their own leg out of bed because they think its someone else's, and they wont see anything wrong with just drawing the right hand side of everyday objects like houses and butterflies. This leads some researchers to comment that it is as if the left hand side of their environment *ceases to exist*

[62] C.f. Bartolomeo et al (1994) and Rode et al (2007).
[63] This is thought to be due to their being some redundancy in processing information for the right hand side in the right hemisphere of the brain, which allows it to compensate and represent both sides if necessary. This is thought to be linked to the fact that in most cases more of the left-brain is devoted to processing language and meaning, and the right hand side is more concerned with spatial processing. I suggest that focusing on left side neglect simplifies discussions without loosing any real explanatory benefit in this context.

(c.f. Becchio 2005: 484). The idea that damage to the SPP stream seems to prevent us from experiencing external objects will become crucial in discussions in Chapter 7, when I talk about representing external objects based on integrating information in the visual buffer with an online egocentric coordinate system. Hence much of what I will discuss here is setting the stage for my final theory, which will draw together all these strands in the final chapter. So in summary: in Visual Neglect it is usually the left side of the visual field that is not attended to or fails to be represented and a similar thing happens in Representational Neglect, where details on the left hand side of the visual images are not reported.

It is worth noting that the term ‘Neglect’ might best be seen as an umbrella term for a collection of mental syndromes that include attentional and representational deficits amongst other things (perhaps in a similar way to autism spectrum disorder). In a recent review of the current opinion on Neglect, Bartolomeo (2007: 384) suggests that Neglect may best be seen as ‘dysfunction of large-scale right-hemisphere networks’, where spatial information involved in attention salience and representation are variously disrupted across a number of modalities. Hence we would expect that different parts of the network can be affected differently depending on the location of the brain lesions, and this links well with Kosslyn’s method of explaining some of the double dissociations observed in this syndrome. Some examples of the range of these dissociations are as follows: in Neglect, not only can visual imagery and vision both be affected together (Bartolomeo et al 1994, Rode et al 2007, Bisiach at al 1979); but also, vision can be spared and imagery affected (Guariglia et al 1993, Ortigue et al 2001, Coslett 1997); or imagery can be spared and only vision affected (Anderson et al 1993, Coslett 1997); and there is even one reported case of right visual neglect and left representational neglect (Beschin et al 2000)[64].

[64] The above list is based on Rode at al (2007: 435). And note that the Guariglia et al (1993) paper reports isolated damage to the right frontal lobe so we can immediately suggest that this explains a top-down problem with imagery instigation, which would not necessarily effect perception.

Whilst investigating and analysing each of these cases in detail would be an interesting project, this unfortunately falls outside the main remit of this thesis; although this is something I would certainly like to continue to do in future[65]. This is especially true since they do not directly threaten the Kosslyn Model, since a similar reply to that given in Section 4.2 can also be used to explain these double dissociations. Specifically we can suggest that, although the visual buffer is intact in each case, we can potentially explain these various deficits by identifying damage to: either mechanisms feeding the visual buffer itself; or to the pathways that help these mechanisms communicate with each other; or to a combination of the two, as discussed above. And since vision and visual imagery use some, but not all, of the same mechanisms and pathways, there is potential for these two abilities to be affected differently or in the same way. Hence in principle the Kosslyn Model can fairly easily accommodate these sorts of results. I will touch on a few of these issues by giving specific examples in the next section, where we will investigate certain ways that the visual buffer could be filled in or accessed in an incomplete way. And we will discuss how this could lead to people only reporting half of the information normally available in an episode of imagery. But for now this completes my introduction and defence of the Kosslyn Model as a whole system. In the next section I will start the job of explaining how this can be seen as a good candidate for an offline simulation process.

4.3 – The Kosslyn Model as an offline simulation process

In this section I will begin to analyse how the whole Kosslyn Model described above, can be seen as an offline simulation process. Some of this analysis will be based on developing Figure 2.2 from Chapter 2, which illustrated a generic offline simulation process (an

[65] A Kosslyn style meta-analysis (c.f. Kosslyn et al 2006 - Appendix) of all these papers and exactly what tasks are being conducted, and where the brain lesions are, would seem a good way of doing this. Some of this has already been done for the background research for this chapter.

updated version is given below). In detail this will involve taking a closer look at a few of the core simulation mechanisms proposed in Section 2.2.3 and seeing how they can be merged or identified with mechanisms in the Kosslyn Model. While previously I introduced seven mechanisms involved in the simulation process, I will not be able to analyse each one in the same amount of detail. Hence I will focus on what I see as the most crucial ones to this context. In any case, certain of the more general mechanisms, such as the Interpretation Capacity and the Executive Controller, can only be given a very broad treatment. This is because they are much more abstract and distributed mechanisms and are hence less well understood (but see Chapter 6 for more on the Interpretation Capacity).

To remind you of the generic offline simulation mechanisms that we might look for in the Kosslyn Model, I repeat Figure 2.2 below, but update it and make it specific to visual imagery. I have labeled some of the mechanisms with suggestions of the mechanism from the Kosslyn Model that seem most likely to fulfill this role. Justifying these designations will occupy most of the rest of this chapter and the remainder of the thesis. In the remainder of this chapter, I will mainly focus on the input (left) side of the diagram and look in detail at how the visual buffer may be filled-in appropriately by internal means in visual imagery. By *internally* I mean from other areas of the brain, as opposed to external stimulus received from light hitting the eyes. This analysis will involve looking mainly at potential candidates for the *Offline Input Identifier Mechanism* and the *Offline Feeder Mechanism*. I will abbreviate these to ‘Identifier Mechanism’ and ‘Feeder Mechanism’ respectively from now on for ease of reference. The remaining main output mechanisms will be discussed in the following chapters, where we will look more at the eventual output from this process in terms of the phenomenal visual image itself and its eventual interpretation. It is worth just mentioning again explicitly that I will assume from now on that the Visual Buffer, as described by the Kosslyn Model, is the equivalent of the Main

Mechanism in the simulation process that is run offline to support visual imagery. Establishing this has been the main purpose of the thesis so far and hence I will not justify this anymore in what follows:

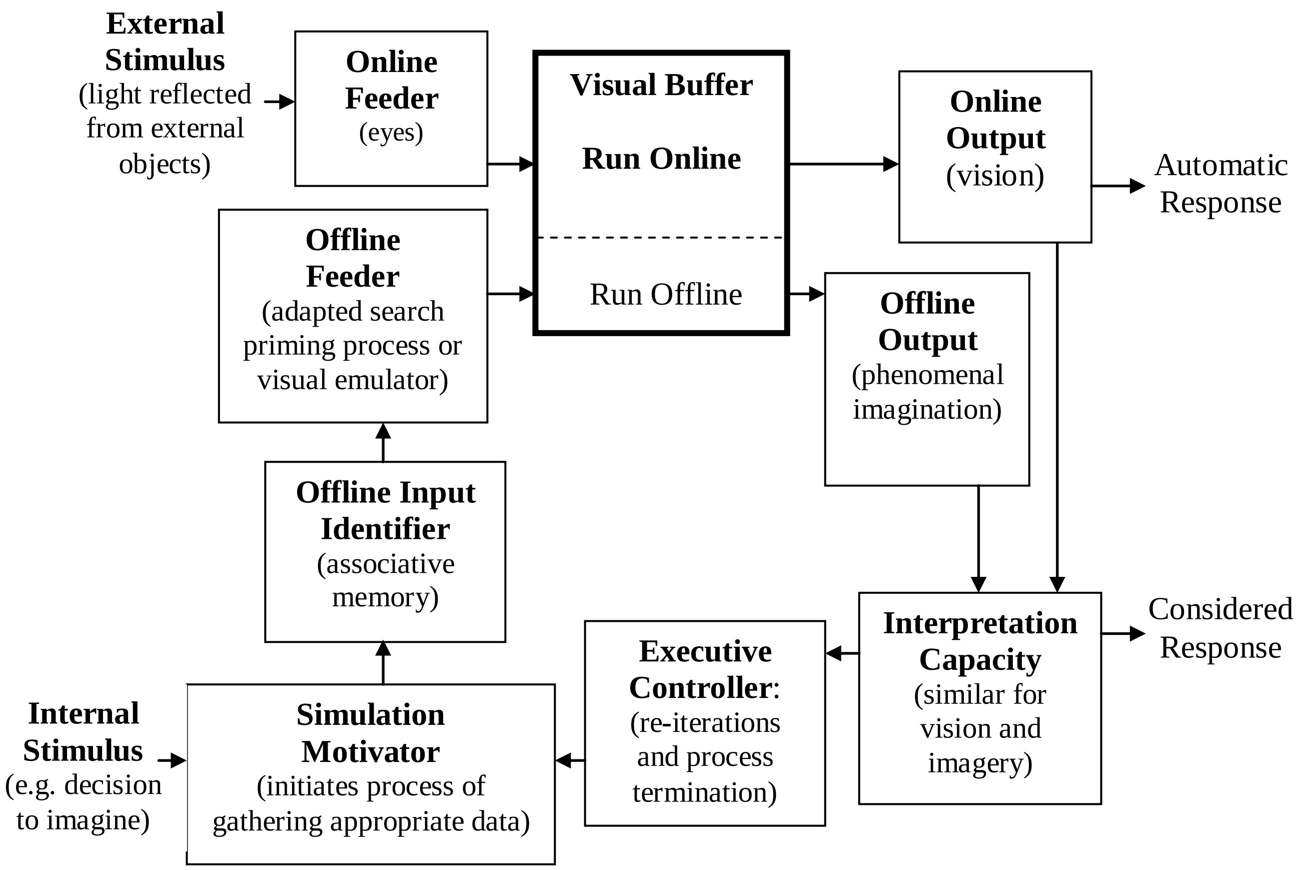


Figure 4.3 – Schematic of the proposed main components involved in full offline simulation involved in willed visual imagery under the Kosslyn Model

As you can see I have repeated all the mechanisms from Figure 2.2 in the diagram above and appended each one with a vision specific suggestion that I will explain in the following. Unfortunately, on the input side (left side), I will also not be able to look in detail at the *Internal Stimulus* and the *Simulation Motivator Mechanism* and only make the following brief comments. By Internal Stimulus I mean the thing that triggers the call for an episode of imagination. This can be as simple as coming to a point in a long thought process which leads you to a realisation that generating a visual image may help solve the problem. This could be fairly obviously consciously willed and deliberate (e.g. the Rubiks

Cube Problem), but it can also happen in a fairly automatic or sub-conscious way[66]. Both of these would count as an Internal Stimulus to generate an episode of visual imagery, and I will focus on obviously consciously willed episodes in this context.

Once this stimulus is generated it sends signals to the Simulation Motivator Mechanism, which is what sets the whole process in motion by activating the other mechanisms necessary for imagery to occur. Its main job is to send signals to the Identifier Mechanism with information of what is to be imagined in order to solve the task at hand. For instance: if it is a Rubiks Cube you need to imagine, then this mechanism will motivate the search for the parameters of a Rubiks Cube to be drawn together; if it is windows in your house you need to count, then this will motivate the search for the parameters to do with the lay out of your house to be gathered together; and so on for other objects that may be required in your imagery process. Hence it is the Simulation Motivator Mechanism that sends the calls out to other mechanisms to start the process and indicates what is required for an appropriate episode of imagery to occur.

Given what I have said about these two very upstream mechanisms, we can perhaps speculate that failure of either of these processes might lead to an inability to generate phenomenal imagery in the first place, even though all the other mechanisms in the system may be intact. It may also be that these mechanisms are under developed in certain people and may explain the variability in our reliance on visual imagery for thinking. For example, this may explain why a small minority of subjects, as discussed in Section 3.4,

---

[66] I appreciate that imagery can initially be stimulated by external sources (e.g. asking someone a problem), but I would still suggest this then requires some *internal* recognition within that problem solving thought process that a visual image may be helpful. At any rate this is the kind of process I will focus on here, hence this doesn't specifically deal with more seemingly spontaneous forms of imagery that may arise in response to, for example, a smell or a sound: e.g. a smell of perfume giving rise to a visualization of the face of your first lover. Again, I think this is an aspect that could be easily dealt with by extending this model but would fall outside the main focus of this project.

fail to experience any imagery at all[67]. Perhaps this is because they: either never really realise they have a need to form it (internal stimulus deficit); or they cannot motivate an episode of it even if they think it will be useful (simulation motivator deficit)[68]. So the above gives us a brief description of what unique roles these two steps on the input side may provide. Given that we can now start to look at some more crucial down stream input mechanisms. We will start by looking at potential candidates for the Identifier Mechanism in the next sub-section, and then look at the Feeder Mechanism in the sub-section after that.

4.3.1 – Offline Input Identifier Mechanism (Identifier Mechanism)

The role of this mechanism is to identify the correct parameters that should be sent to the visual buffer so that they can be depictively represented there and can subsequently lead to the generation of a task appropriate visual image. For instance, once the Simulation Motivator has sent out the calls to gather data together to use in generating an image of a Rubiks Cube, it is this mechanism that actually does the searching for these parameters and makes them accessible. It seems to me that there are two potential ways that the Identifier Mechanism could work within the Kosslyn Model. One involves mainly the OPP stream on its own and the other involves both the OPP and SPP stream. The former would involve imagining an object without particularly needing to locate it anywhere specific spatially, and the latter would involve integrating with the SPP stream in order to delimit what is imagined based on some spatial specifications[69]. It should be noted that the net effect of

---

[67] Recall this was based on a survey asking researchers in this field if they experience imagery and how this affected their interpretation of the Imagery Debate: as discussed in Kosslyn [pg. 179-180].

[68] This way of talking parallels a possible explanation of Autism under an Autism-as-failure-of-imagination approach suggested by Currie (1996:249). This may represent just one way that autistic individuals might have different minds from normal people, in that it may prevent them solving puzzles and imaginatively projecting to other people's perspective. N.B. this was the initial source for suggesting this mechanism in the first place.

[69] And note that in both cases the image will eventually be experienced 'spatially' in terms of it being an apparent 3D object that takes up space. The distinction here is whether it is imagined in a specific spatially located context: e.g. a 3D object in your office, compared with the same object in your bedroom.

each process is to make sure the correct information is accessed from the Associative Memory Mechanism, which is then sent to the Visual Buffer. Hence I have appended the Identifier Mechanism in Figure 4.3 with the Associative Memory label, as I think it is the most likely candidate for this mechanism. We will return to look at each of the two potential processes introduced immediately above later in this sub-section. Before doing that I will briefly introduce a distinction between allocentric and egocentric coordinate systems that will be helpful in the following analysis. This talk of different coordinate systems will also provide a platform for moving on to talk about externally and non-externally represented objects in Section 7.3.2.

i) Allocentric vs. egocentric coordinate systems

As alluded to earlier, egocentric coordinates are literally self-centred coordinates and are those that pertain to a location of objects relative to the subject. Allocentric coordinates, on the other hand, are independent of the location of the subject and refer to the location of objects relative to other objects, or to the relative location of pertinent features within an object. For example, a cube must have corners at equidistant allocentric coordinates by definition, irrespective of *where* it is located relative to you. Consider the following quote from Goodale and Murphy (2000) to illustrate this distinction in coordinate systems further:

> 'To reiterate, allocentric spatial information about the layout of objects in the visual scene is computed by the ventral [OPP] stream mechanisms (which mediate conscious perception), while precise egocentric spatial information about the location of an object in body-centred frame of reference is computed by the dorsal [SPP] stream mechanisms (which mediate the automatic visual control of action).'
> Goodale and Murphy (2000: 192 – square brackets added)

They base their distinction on a subject DF who seems to have relatively little problem reaching for things (egocentric processing), but has problems judging the relative distances between objects (allocentric processing). In tests she can point to a number of counters accurately but struggles when asked to reproduce their relative locations by copying their arrangement with identical counters. This is interpreted as being due to an intact egocentric processing stream, but a dysfunctional allocentric processing stream. The details of this study are relatively un-important here and I just mention them as an example of the principled use of these different coordinate systems in the literature. However it is worth noting that the quote above does in fact also act to indirectly support the visual stream bifurcation used in Kosslyn Model.

The main point of introducing these distinctions here is to illustrate how egocentric coordinates seem more linked to determining the external coordinates of objects, because they determine where the object is in the environment relative to the observer[70]. Where as allocentric coordinates seem more detached from absolute location relative to the observer, because they remain constant across perspectives. This may make them more useful in certain problem solving tasks where the relative shapes and sizes of objects is more important, than actually where they are in the world compared to where you are. To illustrate this, it seems that solving the Rubiks Cube Problem relies more heavily on allocentric coordinates, because it doesn't really matter where the object is relative to you to complete the task, as long as you get the internal geometric relations correct: e.g. the equidistant sides essential to a cube and the way it is divided up into three rows and columns. This is even though of course it may be still be necessary to take a certain 'standard' default egocentric perspective on the imagined situation. Regarding that I would suggest that, broadly speaking, the *default imagery perspective* we employ is just to imagine the object directly in front of you, at some middle distance, at which the object is

[70] There are many ways to define egocentric coordinates e.g. somatocentric (body) vs. cephalocentric (head) vs. retinocentric (eyes) coordinate systems (Mesulam 1999). But since I am interested in the visual field I will be assuming retinocentric coordinates, but refer to the unified single visual field produced from information detected by both eyes.

most commonly encountered from. So unless a specific egocentric location is crucial to the task at hand, then a default one may be used that is probably the one most commonly experienced or is just simple to use and ameneable to the task at hand. I wont say much more about the allocentric coordinate system in what follows. I only introduce it here to offer a contrast to the egocentric coordinate system, which I *will* say much more about in what follows.

So given the above, how does it help us here? Well perhaps imagining a Rubiks Cube is an example of how Associative Memory could be accessed directly in the OPP stream with a minimal demand for accessing the SPP stream. That is a default egocentric perspective is adopted and the Rubiks Cube is imagined using mainly the most familiar parameters stored in memory of its shape, colours and other intra-object details (e.g. texture). In contrast to this we might try imaging the Rubiks Cube in a familiar location from a certain perspective, such as waking up with it at the end of our bed, and try to remember what else we might see from that perspective. In this situation it may be important to access the SPP stream to set the imagined egocentric position and to delimit what might also be seen from there. This is as well as using it to delimit the relative size and shape of the Rubiks Cube in that spatial context. At that point the correct parameters can be sent from the SPP stream to the Associative Memory centres to retrieve the shapes and colours of those objects, which can then be depictively represented in the correct relative locations in the visual buffer. This in turn is what controls what is subsequently visually imagined. In this latter example, merging with the SPP stream in an appropriate manner would be crucial to the appropriate completion of this task. To illustrate this further let’s look at an example of where this integration might go wrong and we may be unable to retrieve the correct depictive information due to damage to the SPP stream.

ii) the Piazza Del Duomo Example

Perhaps the most famous example of where this system can go wrong is from the Representational Neglect literature. For example, in a classic study by Bisiach and Luzzatti (1978), they asked a subject with *associated* Visual and Representational Neglect to describe from memory the Piazza del Duomo when viewed from opposite ends of the square. In both cases the same subject managed to omit features on the left hand side, even though they are included in the description based on the view from the other end. This shows that this disorder is not due to memory loss, but due to an exclusion of details on the left side of the remembered scene from a certain perspective. The simple explanation of this is that their imagery deficit seems to be matching the actual visual deficit they would have, were they to *actually view* the square from these same positions. This seems to indicate that there is something in common in the incomplete way these scenes are experienced by the subject in vision and visual imagery.

In terms of explaining this phenomenon we need to be aware that there is currently a debate about whether Visual and Representational Neglect is mainly due to an inability to *represent* the left hand side of the scene, or an inability to *attend* to it. We can see that both approaches would explain why features on the left hand side are omitted from visual and remembered reports in the Piazza Del Duomo example. If they are represented but not attended to, they will be ignored and not reported. If they are not represented in the first place, then no amount of attending to the left side will reveal the relevant details for reporting. Unfortunately, once again I have to say that, although investigating the debate in this area would be fascinating and instructive, and is something I would ideally like to return to in future, this would also fall outside the main remit of this thesis. It will suffice to say here that there is evidence either way and it seems that both explanations of Neglect have some merit in different situations[71].

The reason we don't need to go into the details here is that the Kosslyn Model can potentially accommodate both approaches to explaining Representational Neglect. This is because we can explain a *representationally based* deficit by suggesting the visual buffer fails to be filled with the information that would normally underpin a visual image of the left hand side of the scene. This could be because only half the normal information is being identified in the SPP stream, which also seems to be having problems processing this information in the online visual case. In terms of the *attention based* explanation, it may be that the visual buffer *is actually* being filled-in completely and in an appropriate manner, it is just that it may not be being accessed fully, or the eventual image that is generated is not being attended to properly. So once again the Kosslyn Model seems flexible enough to accommodate with both these explanations.

The point of introducing this example here was to show what might happen if the SPP stream doesn't interact with the rest of the imagery system correctly. Specifically we may loose the ability to imagine certain scenes with the appropriate spatial properties. So in the process of discussing this deficit we have also described in more detail the second way that the visual buffer could be filled by a combination of the Associative Memory mechanism, the OPP stream and the SPP stream. And along with the OPP stream and Associative Memory only example (default perspective on a scene), this gives us at least two ways that

[71] See Bartolomeo (2007) for a good review of the current thinking in this area. As an example of the use of familiar language: Rode et al (2007:437) discuss evidence that they conclude is: 'far more consistent with a working memory/image generation defect account of representational neglect than they are with a directed attention defect account.'. See also Brewer (1992) for an early philosophical analysis of this field that concludes that one should be wary of trying to oversimplify explanations in this field by reducing them to one single root cause or another (this again parallels earlier warnings about Neglect and those to do with autism spectrum disorder in other parts of the thesis).

the Associative Memory Mechanism could be seen as equivalent to the Identifier Mechanism in the simulation diagram given above. To summarise the two ways discussed here: the first way relied mainly on the OPP stream and made few demands on the SPP stream, perhaps by using only a default egocentric perspective to delimit what was sent to the visual buffer. The second needed to be much more spatially specific and hence drew on the SPP stream to delimit what was present in the scene, and this was then used to guide what information is eventually sent to the visual buffer.

In general terms the identification of Associate Memory with the Identifier Mechanism seems fairly intuitive. This is because if Associative Memory is used online in identifying objects by matching them to certain stored parameters. Then it makes sense to access these same stored parameters to guide how the visual buffer should be filled in the offline case. Therefore I think the above gives a fairly good defence of identifying the Identifier Mechanism with the Associative Memory Mechanism as it is utilised in the Kosslyn Model. It also seems that in the process of defending this match-up, we have also learnt more about how the SPP and OPP streams might interact with the Associative Memory Mechanisms and the Visual Buffer to support imagery that is specific about is spatial context. This marks a major step in explaining at least the spatial part of Spatio-temporal Imagination. I will return to discuss this more in Chapter 6 and we can perhaps now look at the next downstream input mechanism.

4.3.2 – Offline Feeder Mechanism (Feeder Mechanism)

The role of the Feeder Mechanism is to actually feed the relevant information gathered from Associative Memory into the visual buffer. Hence we need to look for examples of mechanisms that seem able to fill-in the visual buffer with depictive information internally. Within the Kosslyn Model, as described above, Kosslyn suggest one possible way that the visual buffer gets filled-in by a method that is similar to a *search priming process*; and I will say a little bit more about that here. However I would also like to briefly introduce the idea that a visual emulator might also be available for the same purpose. While I will not be able to provide an argument that can discriminate which process is more likely to operate, my main purpose here is simply to show that there are at least two options on the table at present; even though the second option will turn out to be slightly less well supported at present. It is however possible that it in the future we will discover that these two processes in fact work together to fulfil this role. The main benefits of both these approaches is that they build up from already existant mechanisms involved in online vision, and hence allow a potentially more parsimonious explanation of their development into a mechanism available for offline processing. This then lends itself to evolutionary style explanations of how they may have emerged, by suggesting they developed for an online purpose in the first place, and then got commandeered for an offline one later. And while I wont elaborate on the ideas behind this approach here, I did make some comments to this effect in Section 2.2.3[72]. We can now look at the two possible candidates for the Feeder Mechanism that I have identified, starting with seeing it as an adapted search priming process.

i) Feeder Mechanism as an Adapted Search Priming Process

As described at the start of the chapter, the Kosslyn Model predicts that the processes normally used for non-canonical object recognition might be similar to those used for image generation. This is where information from the Associative Memory areas are broadcast to other areas (e.g. Information Shunting and Attention Shifting), so that useful features can be searched for in the visual scene. Kosslyn calls this *search priming* and this

[72] Recall I quoted Currie and Ravenscroft (1997: 177) and Currie (1995a) suggesting that it may be a good evolutionary trick to commandeer already existant mechanisms in a ‘parasitic’ way, because this leads to a more parsimonious explanation of how these offline mechanisms could develop, all other things being equal.

enables a slow, more involved method, of object recognition. They explain how this relates to generating imagery in the following quote:

> 'According to our theory, the same mechanism that is involved in priming the stored representation of a distinctive part that one expects to see during perception is also used to prime the representation during imagery. In imagery, however, the priming is strong enough that the representation is evoked in early visual cortex, reconstructing the shape of the object in the visual buffer.' Kosslyn et al (2006: 155)

They also refer to Kosslyn (1994) for details of the neural implementation of this process[73]. From these two sources we can gather that although these two processes are similar, there is one crucial difference. That is, in non-canonical object recognition the primed areas do not go so far as filling-in the visual buffer, they just alert other areas to search for certain things within it. In other words, they are primed to check for certain expected patterns in the visual buffer.

In terms of imagery on the other hand, the extended claim from Kosslyn is that in image generation this process is to a certain degree more active. So much so that in the imagery version of the priming process it starts to actually affect the contents of the visual buffer to such an extent that it actually starts to insert depictive contents into it. These internally generated contents then somehow form the basis for the eventual vision-like image we experience. Maybe we can describe the transition from one process to other like this: we are looking so hard for a certain feature that in the end we start to generate an image of it in our minds. That is, the priming is so strong that the features we are looking for, somehow actually get inserted into the visual buffer, and this is the basis for the eventually experienced visual image. In terms of our eventual capacity for fully willed imagery, it seems that at some point this ability would have to have come under conscious control so

[73] And this feature is particularly discussed and built up in the last three chapters of that book.

that, instead of just searching for objects and imagining them, we eventually could just imagine any object that we wanted to irrespective of the current visual context. So perhaps it is this that formed the basis of a separate visual imagery capacity that we can now completely separate from a search priming process. At this point an *adapted* search priming process can be accessed at will to fill-in the visual buffer with information relevant to the task, as delimited by upstream processes, and as discussed above.

Perhaps then this suggests one way that data stored into the Associative Memories is used in a top-down (retrograde) manner to generate information that the Feeder Mechanism inserts into the visual buffer. To a certain extent this is a reversal of the normal object recognition process: where instead of trying to recognise objects represented by data in the visual buffer which is fed upstream; we are using stored information of previously recognised objects, so we can send them back downstream to feed into the visual buffer. However this data, which has been deliberately sent downstream, is not experienced as a normal visual experience, it is experienced as a pseudo-visual experience with apparent shapes and colours. So I would suggest that somehow the data coming downstream is tagged or labelled differently so that it is not treated as an online input and hence not normally experienced as an external object. Hence this information coming from the Associative Memory is somehow tagged appropriately as mock data and is treated as such by other mechanisms that then access the visual buffer. Making sure this information is stamped as ‘mock’ before feeding it into the visual buffer, could be one of the jobs the Feeder Mechanism does as it fills it in. The Feeder Mechanism could either do the job of actually stamping this tag on or merely double check that it is there. That mock label is therefore what ensures that the information it feeds in is not at risk of being taken seriously as an external input by ensuring it is tagged appropriately. Perhaps the search priming process already makes this distinction between information involved in a search as compared with the data that is representing external objects. Which then may mean this step also has a usefully commandeered pre-cursor. And again we can speculate that a breakdown in this checking or stamping process might lead to internally generated but externally experienced pseudo-visual events: e.g. hallucinations.

Putting it this way then allows us to see how this can easily be interpreted as a simulation process where: stored data is drawn from Associative Memory and sent to the Feeder Mechanism; this is then tagged as mock, and hence identified as an offline input, which is then fed into the visual buffer by something close to an adapted search priming process; the visual buffer could then be accessed by other mechanisms to generate the finally experienced visual image. So as long as the search priming process can be adapted to feed willed inputs into the visual buffer, this certainly seems to give us one good candidate for the Feeder Mechanism. Although establishing this further would take significantly more work, this seems at least in principle to provide one possible candidate for the Feeder Mechanism. But simply identifying a possible candidate is all that was aimed for at this point, so we can now perhaps look at the second possible candidate, which is based on visual emulators.

ii) Feeder Mechanism as an Adapted Visual Emulator

In this sub-section we will look to see what evidence there is for visual emulators in the first place and then discuss how this may be potentially employed as a Feeder Mechanism. Given the distinctions made in Section 2.3 between Emulators and Simulators generally, a visual emulator would strictly only be involved in improving, or fine tuning and smoothing, the performance of the currently active online visual system. This might involve anticipating certain features of the scene that one is about to observe and preparing the visual system to process these. Alternatively it could help by filling in sensory data where the original data is degraded. This is certainly something that Kosslyn and Sussman (1995: 1035) suggest happens, where they argue that imagery: 'is used to complete fragmented perceptual input' and to 'prime the perceptual system when one expects to see a specific object'. They offer some evidence of this in terms of a phenomenon that is known as 'representational momentum'. Here one falsely remembers that an object's finishing point was: 'farther along in the trajectory than it actually was' [pg. 1039]. Hence showing some anticipation of an objects future position and the confusion of this with its actual finishing point. This gives us one way of suggesting that feeding the visual buffer internal inputs was initially developed to aid online vision by predicting future states and filling-in areas where externally inputted information was degraded or incomplete.

In addition to the above, Grush (2004: 392) quotes evidence from Duhamel et al (1992) that suggests: 'there are neurons in the parietal cortex of the monkey that remap their retinal receptive fields in such a way as to anticipate immanent stimulation'. Providing evidence that the visual areas are anticipating data that they may be about to receive. He is very explicit about this interpretation in a later paper:

> '…the PPC [posterior parietal cortex] neuron will begin firing shortly after the motor command to move the eye is issued, but before the eye has actually moved. The PPC neuron appears to be anticipating its future activity as a function of the current retinal projection and the just-issued motor command. That is, it is a visual modal emulator.' Grush (2007: 401)

So he clearly counts this as fairly good evidence for a visual emulator working in a single sensory mode[74]. Grush (2007: 400) also thinks that evidence from a study by Rao et al (2001) supports the idea of a visual emulator. This because that study suggests that the

[74] This distinguishes it from amodal emulators that seem to 'emulate' the environment using information from multiple-sensory modes, in an amodal way. However this now sounds like simulation since it is not fine-tuning any online performance, but rather helping to strategy test. Thus illustrating another way my distinctions in Chapter 2 might be applied.

visual system not only uses information from events that have just happened to improve current visual accuracy, but they also use information predicted from future events to do so as well. This again re-enforces the idea that there may be some mechanism anticipating likely future visual states and using these to improve performance in the present. Much of the other evidence he cites in his 2004 article, which was analysed extensively in Section 2.3, I would interpret as talking about simulation processes, rather than emulators as such. So he does not provide as extensive a defence of an emulation process in the visual modality, as compared with motor emulators. Which is natural given his background focus on motor control; hopefully more details will be forthcoming from Grush in the visual arena in the future.

So in summary, I would suggest that current evidence for visual emulators is suggestive rather than compelling, but there is certainly some potential for developing this approach in the future. However, even if the existence of a full visual emulator was eventually confirmed, there is at least one important disanalogy between it and a motor emulator. And this is that the visual emulator seems to be already filling-in the visual buffer directly in order to improve online vision. This is in contrast to the motor emulator, which was seen as a separate mechanism that only helped fine-tune the signals sent from the Motor Controller to the body. Hence the motor emulator only ever interacted with the Motor Controller and never sent any commands directly to the online system (i.e. the body). However this disanalogy makes the visual emulator even more ameneable for commandeering as an offline simulation process. This is because if it is already feeding inputs directly to the visual buffer, this is just what we want to use it for as a Feeder Mechanism. Except in this case we are not feeing it inputs for what we might expect to see in our vision in the near future, we are feeding it parameters of things we would want to imagine. If this change of input is possible, then it seems that at least in principle, a visual emulator could be commandeered as an *internal* way to feed appropriate information from associative memory into the visual buffer.

The above perhaps offers another, admittedly more speculative, way to internally fill-in the visual buffer. Unfortunately I can't really go into any further detail here, because until more is known about how the actual online visual emulator is thought to operate, it is hard to predict what specific features it offers to the offline simulationist. Hence the aim here is merely to point to visual emulators as a potentially useful starting point for developing ideas about this mechanism further; and again this is something I would like to peruse in the future. For now, and in line with the aims of this sub-section, I offer this analysis merely as a way of identifying another potential candidate for the Feeder Mechanism.

So as a way of drawing this section to a close, we can perhaps suggest that if any of these lines of evidence are at least initially plausible, then it could lead to the following three-way split in how information can get sent to the visual buffer. These are: 1) external and online from the eyes; 2) internal and online from a kind of visual emulator; and 3) internal but offline from the associative memory. It may be useful to distinguish these in a short hand form by returning to elaborate on the notation utilised in the last chapter, as follows:

1) Visually **Perceived**: {**online, external**} Array Cell (X, Y): [colour (RGB: P, Q, R), depth Z, etc]
2) Visually **Perceived**: {**online, internal**} Array Cell (X, Y): [colour (RGB: P, Q, R), depth Z, etc]

3) Visually **Imagined**:{**offline, internal**}Array Cell (X, Y):[**mock**, colour (RGB: P, Q, R), depth Z, etc]

In the above I have appended the cell-by-cell contents of the visual buffer with a prefix to indicate whether they were filled-in from internal or external sources and whether they are 'mock' entries or not (see bold text). It is also worth noting that of the three entries, the second has also been given the least support at present, since the evidence for visual emulators is still accumulating. Fortunately, this is the one we are least interested in from now on, since we will mostly be focusing on the differences between the first and the third entries in what follows. The second entry was introduced here as a suggestive way that the internal filling-in of the visual buffer may have been aided by the presence of a visual emulator. Now that has been discussed, the second line will not be referred to again and hence will not need developing or defending further at present.

So by now focusing on the first and third lines perhaps we can suggest the following. The first ‘**Perceived’** entry would be treated as an online real input since it has been ultimately caused by external sources. This is what supports normal perception of external objects once the visual buffer is accessed appropriately. The exact nature of this process, and how it relates to visual imagery will be the subject of the rest of the thesis. For now we can say that we have a good theory of how the visual buffer may underpin online visual perception. The ‘**Imagined’** entry, on the other hand, is appended with the mock label so that it is prevented from being treated as a real external input. Although I have said little about this, this is assumed to occur somehow in the processes that input data into the visual buffer internally. I have suggested that this might be a job the Feeder Mechanism might usefully fulfil since it is the last step before mock data is entered into the visual buffer. This means that information about imagined or remembered objects can be filled into the visual buffer internally without them interfering with our normal visual processing. It also means that when this information is accessed from the visual buffer as an output, these mock entries can be treated differently and potentially kept in a separate imagery stream. How separate this is and how much it is like the online process, is the subject of the second half of the thesis, where we look at what may happen next on the output side of the imagery process. For now I feel I have achieved the main goal of this sub-section, by identifying a few lines of evidence that indicate how the visual buffer can be filled-in internally. This therefore gives us a few options of how an Feeder Mechanism can be explained in terms of the Kosslyn Model and how this can also be seen as an offline simulation process. This concludes my analysis of the 'input' side of the diagram in Figure 4.3. We will look at how this may underpin full phenomenal imagination in the next chapter.

So in summary in this chapter there have been two main themes. The first was to introduce and defend the Kosslyn Model as a whole. I did this by introducing some other mechanisms and streams that are thought to work together with the visual buffer in order to underpin visual imagery. To illustrate how this system works, I went through some positive evidence in favour of it and explained how it can defend itself against some putative counter examples from the brain damage literature. This illustrated the complex and flexible nature of the Kosslyn Model and hopefully showed how robust it is and why I have chosen to adopt it in what follows.

The second theme was to start to look at the Kosslyn Model through the eyes of an Offline Simulation Theorist and to start to show how features of that model could be seen as the equivalents of offline simulation mechanisms. I have particularly discussed the Feeder Mechanism and the Identifier Mechanism in detail here, and suggested which parts of the Kosslyn Model they may integrate with. Specifically I have suggested the Identifier Mechanism can be identified with the Associative Memory Mechanism and the Feeder Mechanism can be identified with an adapted search priming process or an offline visual emulator. I have also discussed the full input stream in general, as a way of explaining how an episode of STIm is instigated and directed in the first place. We have particularly seen how relevant information about the spatial context (i.e. where you are imagining) could be accessed and used to control the information sent to the visual buffer for depictive representation. This has given us an idea of how spatially specific visual imagery might be generated and this has given us a good explanation of the spatial aspect of STIm. I will discuss the more temporal aspect of STIm in Chapter 6, where we will also look at how these visual images are interpreted. The net effect so far has hopefully been to show how the Kosslyn Model seems like a good candidate to treat as an offline simulation process.

What remains to be done is to discuss the more output side of Figure 4.3 by looking at how the contents of the visual buffer might be accessed and transformed into a phenomenally experienced visual image. This marks the transition from looking at the sub-personal processes under-pinning visual imagery, to looking more at the philosophical theories of perception that might have something to say about phenomenal imagination as an offline simulation. Hence in the next chapter we will start to look at how adopting a S*elective* versus *Generative Approaches* to perception might affect the debate at this point.

PART II – STIm and Phenomenal Imagination

Selective vs. Generative Approaches to Perception and Visual Imagery

Introduction to Part II

In the last chapter we looked at the Kosslyn Model as a whole system and analysed whether it could be seen as an offline simulation process. We focused there mainly on the input aspects of the simulation process that would feed different kinds of information into the visual buffer. I concluded that it looks like the Kosslyn Model is quite amenable to being interpreted as an offline simulation process and identified many of the mechanisms we would expect to find in such a system. So far particular attention has been paid to the following: the Main Mechanisms that we would take offline (the Visual Buffer); the Identifier Mechanism (Associative Memory); and the Feeder Mechanism (as an adapted search priming or visual emulation process). I had also described how the visual buffer can get filled with hybrid depictive representations that represent: either a real external object in the case of online perception; or an imagined object in the case of offline visual imagery, as long as they are labelled as mock entries. Additionally in the first half of the thesis we have been mainly looking at the sub-personal processes that might underpin phenomenal imagination and have particularly identified a way that spatially specific information could be gathered and represented in the visual buffer by integrating it with the SPP stream. This has given us a way of describing the underlying way that the *spatial layout* of STIm could be controlled and I think this therefore represents a solid platform from which to continue the thesis.

In Part II of the thesis we will start to look at the phenomenally experienced image *itself* and explore some ways that this could be analysed based on what has gone before. So what we require now is some analysis of how the real (online) and mock (offline) entries in the visual buffer get accessed and can potentially lead to these different styles of phenomenal experiences. This represents a shift in focus towards firstly, the output side of this process (c.f. right hand side of Figure 4.3), and secondly, to analysing the more personal level phenomenally conscious experiences themselves. The next three chapters will explore this aspect by building on what has gone before.

For me, the main issue to clarify at this point is to understand *what might be happening in the online visual case*, so that we can then see how closely this can be applied to the offline imagistic case. This naturally leads to an analysis of certain perceptual theories of phenomenal experience that can help explore this idea. To my mind, the most important and useful distinction to utilise at this point is between *Selective* and *Generative* Approaches to perception. These terms were introduced briefly in Section 1.4 and I attributed them as originally coming from Robinson (1994: 70). However I also follow Fish (2009: 137) in how I use these terms as in the quote below, and have generally found this way of talking is optimal for helping clarify the crucial issues discussed here:

> ‘According to generative accounts, the relevant neural processing generates a phenomenal character for the mental state; according to the selective account, “the process does not generate a content but puts one in touch with the stimulus that is already out there” (Robinson 1994: 70)… This [selective] conception of the role of the brain in experiences sees it not as the organ that somehow generates conscious experience, but rather as the organ that enables us to detect different features of the world in which we live.’ Fish (2009: 137)

So in what follows I will look at what I think this crucial difference means to the imagination-as-simulation debate and these terms will be explained further below.

I have also illustrated how initially I see these two contrasting approaches as differing in terms of what they say about visual imagery in Figure 1.5, where I illustrated how the same phenomenal generation mechanism might be used both in perception and imagery in a more *Conjunctivist* approach. That is they seem to use most of the same mechanisms, including one that generates phenomenal character, and hence there is much overlap or conjunction, between the kinds of things they are. This is as opposed to a more Disjunctivist Position that might deny any common overlap at a certain level of description[75]. This is because they want to preserve veridical perception as something that puts one directly in touch with external reality, and things like hallucinations would be classed as a different kind of thing. This is summarized in a disjunct like the following: EITHER you are having a veridical experience as one kind of thing; OR you are having an experience, which is another kind of thing (e.g. hallucinations). These distinctions will be explained further in Chapter 5.

It is the combination of a Generative Approach and a Conjunctivist Position that I will eventually adopt and my main purpose here is to justify taking this route. A Generative Approach, as I will now develop it here, suggests that the brain can *generate* the phenomenal character of both perception and phenomenal imagery in fairly similar ways, with obvious differences to account for the difference in phenomenology. The opposite of this, in my opinion, would be a Selective Approach, which suggests that, at least in the perceptual case, the brain and body (e.g. sense organs) merely *selects* what external features you become aware of, but doesn't then also generate the phenomenal character itself. In this sense the brain merely facilitates you becoming aware of the qualities that are already out there: e.g. the colours and shapes of external objects. I think this sort of approach is more compatible with a Disjunctivist Position, which might deny any common

[75] And recall I have followed Johnston's (2004) terminology here, but it is worth noting that he tries to steer a path between a Disjunctive and Conjunctive Approach. As I see it he wants to avoid Conjunctivism because of 'veil of perception' concerns, and wants to avoid Disjunctivism because he finds it lacks explanatory power on the non-veridical side of the disjunct (see Chapter 5). Unfortunately I won't have time to deal with his work in detail here, but simply use this terminology in a similar way to him.

overlap between visual perception and visual imagery, at least in certain crucial ways that will be discussed. These two approaches will be described in turn in more detail in the following chapters, but as a way of motivating taking this line of analysis, I will give a further brief justification here.

As already mentioned, by the end of the thesis I will adopt a **Generative Approach**, which accepts that the brain does generate phenomenal character[76] and it is versions of this that we are aware of in all cases of phenomenal consciousness. This is the general claim but I will focus mainly on the visual modality in this context. A Generative Approach is in line with theories that accept that phenomenal character supervenes locally, as illustrated in the Brain in a Vat (BIV) thought experiment[77]. This position therefore accepts that a micro-physically identical brain to yours that is being nutritionally sustained and fed fake inputs identical to your perceptual inputs (i.e. a BIV), would also be having an identical experience to yours. To put it another way, in this situation you could not tell the difference between being you, or being a BIV. This is explained because if you were this BIV, you would also be enjoying the same full range of vivid phenomenal experiences. Because of this I will treat this approach as accepting that phenomenal character supervenes locally (or narrowly) on the brain and this also gives the first indication of why it can be seen as a highly Conjunctive Position to hold. That is, all visual phenomenal experiences are generated in a fairly similar locally supervenient way, it is just the way that they are represented and experienced that differs. This latter point will be elaborated on in detail in Chapter 7.

---

[76] And recall from the introduction that I use the term 'phenomenal character' non-technically to merely indicate the qualitative character of your phenomenal experience.

[77] Local supervenience can also be expressed as: if mental states locally supervene on brain states, then being in the same brain state entails being in the same mental state. This can be paraphrased by saying if two persons are indistinguishable in all of their physical properties, they must also be indistinguishable in all of their mental properties (c.f. McLaughlin & Bennett 2005). I use it here in the restricted context of discussing phenomenal character as indicated in the text. For BIV see: Putnam (1981) and Brueckner (2004).

The benefit of adopting this kind of approach in this context is as follows. If the brain generates phenomenal character in the case of online perception, then this seems like a process that could also possibly be commandeered to generate the phenomenal character of offline visual imagination. Of course these processes will have to be different at some point to explain their different phenomenology, but this kind of theory predicts these modifications will be minor compared with the differences predicted by non Generative Approaches. These minor modifications that might need explaining include why we have experiences of objects as being external and in the world even though they are locally supervenient mental states, and why we don't experience phenomenal visualizations in the same way. Hence some explanation is owed as to how this difference in phenomenology is supported. This dept will be paid in Chapter 7, where I will use the difference between a real online input and a mock offline input to underpin a difference in how these states are represented in our experience. Hence for me, this seems like a promising route to take in order to explain these phenomenal experiences and how they relate to each other, and is the main reason I have taken this expository route here.

But before I can develop that positive account, I think I need to explain why I reject Selective Approaches. Explaining what an example of this approach entails and why I don't favour it, is the main focus of Chapter 5. There, apart from introducing a paradigmatic example of Selective Approach and rejecting it, I will also offer one possible way that phenomenal imagination could be accommodated within this approach. I have already indicated what form this might take in Figure 1.5 and this diagram will be repeated and updated in the next chapter. However I will not spend much time exploring and developing this formulation, as I will ultimately reject it on other grounds.

A **Selective Approach**, as I develop it here, is one where the brain does not generate phenomenal character itself but rather puts you in touch with what's already out there. That is you are directly aware of the phenomenal properties of the object itself and this is what supports the phenomenal character of your experience. This means there is no need for the brain to generate any phenomenal experiences in perceptual cases, since simply becoming aware of the object has the same effect phenomenally. This is why we might be able to regard this kind of approach as being widely supervenient on the object, the brain and all the causal processes in between (e.g. light propagation). In this sense all your brain and body does (e.g. your eyes), is *select* which range of electromagnetic waves you detect (e.g. the visible spectrum) and then processes these, in order to somehow make you directly aware of what is already out there. This is sometimes referred to as 'your mind *reaching out* to put you in direct perceptual contact with external objects', and it is this claim that I will put pressure on in terms of the Time-spread Argument in Section 5.3.

In order to keep this discussion focused I will be forced to concentrate on one paradigmatic example of a Selective Approach known as Naïve Realism. This theory is gaining in popularity in the recent literature and is very closely linked to Disjunctivist Positions. This will be explained below but initially shows why it is a good one to choose, as it seems to make fairly contrary claims to mine by denying any common overlap between vision and other vision-like states (e.g. phenomenal hallucinations and imaginations). Hence rejecting it will also act as way of motivating the Generative Approach that I will adopt later. I will say a few words at the end of the Chapter 5 about how my arguments against Naïve Realism might generalize to other Selective Approaches that would claim perceptual phenomenal character has a wide supervenience base.

It should be noted that Selective Approaches are very closely aligned with Relational and Direct Realist theories of perception, which claim that in a perceptual relation the object you are aware of must exist for you to be aware of it. The sort of perceptual relation you are in is one of acquainting you *directly* with the object as it is and hence necessarily involves the object itself. This is contrasted with how a Representationalist might construe the situation where it is the representation that is the more crucial aspect. Crane (2006) discusses these distinctions at length and I introduce these terms and use them according to his definitions. Naive Realism is a very obvious example of a Relational perceptual theory, although there are others that are very closely aligned to it. For example 'the Theory of Appearing' defended recently by Alston (1999) and Langsam (1997) is explicitly a relational theory and can be treated in a very similar way to Naïve Realism. As mentioned above I will focus on Naïve Realism here, although I think my critique does extend to these theories also (see Chapter 5 for more). I will introduce and analyse more Representational-style theories in Chapters 6 and 7.

I will ultimately reject Naive Realism by using, what I have termed, 'the Time-spread Argument', which is an extension of the more classical Time-lag Argument. This will basically suggest that the external world is not actually in the right configuration anymore for you to 'reach out to it' by the time the light reaches your eyes, and hence it cannot play the role of constitutively being part of your experience. In other words, if the properties your mind is reaching out to have minutely changed by the time your perceptual state occurs, then it seems a mystery to me what your mind reaches out to and how it does it. And further due to differential travel times from objects to your eyes, the scene you are eventually aware of is actually composed of different features spread backwards through time, and hence this puts pressure on the direct realist claim that you see objects 'as they were in the past' (c.f. Le Morvan 2004: 224).

The Time-spread Argument will be discussed in detail in Section 5.3, and will show how appealing to the visual buffer within the Kosslyn Model might help illustrate these points further. This is because under this model the details we are eventually aware of in perception seem delimited to the current contents of the visual buffer and not the configuration of external objects as they *were* spread backwards through the past. This use of the visual buffer to comment on perceptual theories represents a reversal in its normal role in my arguments up to now. Before I was concerned with defending its role in visual perception and visual imagery, but perhaps now that I have defended and accepted it, I can use it to explore ways that it might guide our choice of perceptual theories.

As mentioned already, I will briefly analyse what a theory of visual imagination based on the Kosslyn Model might look like under a Selective Approach. The general idea here being that this way of describing perceptual states seems much less ameneable to being commandeered in an offline simulation process. This is because if in the online perceptual case the information in the visual buffer is supposed to support you getting in direct perceptual contact with the object, it seems that this 'reaching out' function (if I can call it that) is not something that is usefully commandeered for visual imagination. This is because in visual phenomenal imagination there is no equivalent object in the world, or indeed in the mind, that we could reach out to and therefore it seems we are left without an equivalent object whose properties we can 'latch' on to. Hence I will suggest that a separate phenomenal generation step is required for phenomenal imagery to occur and this means that the mock contents of the visual buffer could be treated very differently to the real online ones when they are processed upstream.

So if one adopted a Selective Approach to perception, this suggests to me that the visual buffer will be accessed in very different ways depending on whether you are dealing with online versus offline entries. This therefore seems to represent a divergence in the kinds of mechanism these processes might use further upstream, and hence this might be an end to where we can find online mechanisms suitable for taking offline. In a nutshell, I am saying that I don't think the mechanisms supporting the output side of selective-style online perception will be suitable for taking offline in order to generate offline phenomenal imagery. And hence this could represent an end to the offline simulation process and the start of another process that uses mock inputs in markedly different way to perception. If this is plausible, then to me this also makes the disjunct between the two processes very clear, in that you are EITHER: selectively reaching out and directly visually perceiving a currently present object; OR, in other vision-like cases like phenomenal imagery and hallucination, you are locally generating a vision-like experience of a non currently occurring object. This is the premise of the second half of the thesis and why I think it is important to spend some time exploring Selective and Generative Approaches in order to develop a full theory of phenomenal imagination and hence STIm. But before looking at a Selective Approaches in Chapter 5 in detail, I will make a few more clarifying points regarding the second half of the thesis.

Up to now I have been talking as if the visual buffer is 'accessed by other mechanisms', which then go on to form perceptual and imagistic states. However it is quite possible that the visual buffer actively *sends* its information somewhere, or can even support phenomenal perception or imagery on its own. Hence I have implicitly used the phrase 'accessed by other mechanisms' to stand in for a more unwieldy sentence like: 'the visual buffer is accessed by, or sends its information to, another mechanism, or then goes on to process these further itself'. So at this point I am explicitly mentioning that I only use this phrase for ease of reference to represent the idea that this information moves upstream in some way and has further processing applied to it before the final phenomenal product is realised. I will say a little bit more about what I think this might involve in Section 7.3.1, where I look at theories of consciousness that cover a range of ideas as to what else is necessary for phenomenal experience to occur.

Also before we go on it is worth pointing out that, although I focus on phenomenal imagination here, this does not preclude the possibility that some mental imagery processing occurs at a more sub-personal level. This point is illustrated by noting that the Appendix of Tye's (1991) book called 'the Imagery Debate', is composed of a computer programme that iteratively 'rotates an object' in 45 degrees increments and compares this to a 'target object'. The idea being that this is a symbol manipulation equivalent to mentally rotating a phenomenal image of an object until you can tell if it is a rotated version of a target object. But obviously this is a computer programme and hence this is occurring in a purely mathematical and algorithmic way by manipulating binary symbols. The equivalent versions of imagistic rotating and scanning experiments were discussed briefly in Section 3.2. The point I making here is that an equivalent more sub-personal kind of processing may also occur in the brain, based on fairly mechanistic neuronal firings, without us being consciously aware of it. And by focusing mainly on phenomenal imagery I am not denying this point. It is only that I am deliberately restricting my attention to phenomenal imagery, which I find more interesting and I think is more appropriately approached from a philosophical perspective. Hence my focus here is not prohibitive of sub-personal imagistic processing that is analogous to the computer programme mentioned above. It is merely restricted to looking at personal level phenomenal imagination as a way of limiting the scope of the thesis.

Finally, in terms of doing the groundwork for the second half of the thesis, in Section 2.2 I lumped together the Output Detector Mechanism and the Offline Output. Since we are now looking at the output side of Figure 4.3 in more detail here, I will just briefly explain this move again. The Output Detector Mechanism is thought to be the mechanism that differentiates between online and offline entries in the visual buffer and would potentially siphon them off for upstream processing according to which category they fall into. This will ensure they get treated differently and we don't end up confusing what we have imagined with what is really out there. I wont say much about this in what follows and merely assume it occurs in some way. And although this may seem like an inefficient extra process to include, it need not be interpreted in this way. Consider the sieve analogy proposed by Currie and Ravenscroft (2002: 70), where it is simply the size of the stones that make the difference as to whether they follow one path or another, rather than needing any extra sorting effort to occur. In a similar way, simply having the mock tag attached, may be enough for the offline entries in the visual buffer to be treated differently by the processes accessing them. Hence in what follows I will assume this Output Detection and differentiation occurs fairly automatically and potentially fairly easily in our mental economy[78]. And we will therefore focus mainly on the Offline Output and I this will be taken to mean an episode of phenomenal imagination in this context.

There is also one other output mechanism that I will discuss in a bit more detail in Part II of the thesis, and that is the Interpretation Capacity (see Figure 4.3), which allows us to interpret the visual image we are having in a useful way relative to the task at hand. We will look at this in Chapter 6, by analysing an idea proposed by Tye (1991: 94) about how visual images are labelled and interpreted. However I will modify and expand his account

[78] Although of course we could suggest that even if this sifting process does incur a significant extra efficiency cost, this might be worth it given the predictive powers and problem solving abilities it may allow. So as long as the benefits outweigh the costs then this may not be a preventative necessary extra step.

to one that I think reflects our experience of imagery more closely and also deals with a wider range of the possibilities in this area. It is at this point that I will start to introduce how a visual image can come to be interpreted as representing different *temporal* and *spatial* contexts, even if the *spatial layout* of the objects in the image remains the same. Hence giving us a suggestion as to how spatio-temporal aspect of Spatio-Temporal Imagination (or STIm) is supported. Before dealing with this aspect of Tye's work, I will also briefly explain why I don't adopt his form of Representationalism about perception and phenomenal imagination. This will involve looking at some problems for this kind of theory and showing how I think it may be rejected along with other Selective Approaches.

So having said all that, we can now progress to look at Selective Approaches in more detail, and more specifically we will look at a paradigmatic example of this which is known as Naïve Realism.

5 – STIm and a Selective Approaches to Perception:
Naïve Realism and imagination-as-simulation

> 'I simply do not understand how material objects, understood in a realist way, can be literally parts of experiences. And to say, as is sometimes said, that "the mind spreads itself on the world" only adds to the obscurity… and most important, even if the view in question were intelligible and known to be correct, it would still remain true that there would be no difference between a veridical experience and a suitable hallucinatory experience that would be genuinely discernible to the perceiver in question -and so no genuinely internalist reason for thinking that the belief in question is true.'
>
> Bonjour (2004: 363 – footnote 32 –addressing work by Brewer 1999)

In this chapter I will introduce and analyse Selective Approaches to perception by looking at a specific clear example of this approach called Naïve Realism. Along the way I will also explore its implications for the imagination-as-simulation approach as I see it. So specifically, I will try to do the following three things. **Firstly**, I will introduce and explain Naïve Realism, which I will take as my paradigmatic example of a Selective Account. This will give us an idea of the things this theory claims so that I can then criticise it later. I will then look at how this theory defends itself against the Argument from Hallucination and how this may relate to some arguments about the overlap between perception and visual imagery. This will explain why naïve realists are usually also Disjunctivists, and why, conversely, I consider my theory to be a highly Conjunctive one, with much overlap between perception and phenomenal imagination. At this point I will also suggest one way that Selective Theories could accommodate locally (or narrowly) supervenient phenomenal imagination, even if they adopt a widely supervenient relational theory in terms of perception. **Secondly**, I will look at two motivating factors for defending Naïve Realism and explain why I don't find them very compelling. This will hopefully explain why I find issues such as the Time-lag Argument more crucial in guiding my choice of perceptual theory. **Finally**, I will explain in detail why I don't favour a Selective Approach like Naive

Realism, based on an extended version of the Time–lag Argument, which I call the Time-spread Argument. I will also explain how the work done previously on the features of the Visual Buffer might help guide us here. This will allow me to develop a positive Conjunctivist account of phenomenal imagination in Chapter 7 based on a Generative Approach. I will start by introducing and analysing Naïve Realism in a bit more detail next.

5.1 – Naïve Realism as a Selective Approach to Perception

For something to count as a Selective Approach according to my usage of the term, it mustn't suggest that in veridical perception the brain generates the phenomenal character. Instead the role of the brain is merely to help *select* what one is aware of in the external world. Hence the phenomenal character must in some way depend on some *perceptual relation* between the observer and the object and the object must play a constitutive role in our awareness of it (cf. Crane 2006). To my mind modern accounts of Naïve Realism seem to be paradigmatic examples of this kind of approach, and therefore I will introduce and analyse a version of that here. The version I will analyse has been defended recently by researchers like Martin (2004, 2006), Fish (2008, 2009) and Kennedy (2007, 2009), and as mentioned already, Fish (2009: 137) specifically endorses and discusses the utility of categorizing Naïve Realism as a Selective Approach.

It seems that defenders of a position called 'the Theory of Appearing', also seem at least sympathetic to this kind of approach, for instance Langsam (1997: 53) specifically endorses this association to Naïve Realism and Alston (1999: 183) may be sympathetic to the more modern version of it as discussed here. It certainly seems to bear the same crucial features and can perhaps therefore be a target for this critique in a fairly similar way. I will only discuss Naive Realism explicitly in what follows, but this extension to the Theory of

Appearing will be implicit in everything I say in this chapter. I will say a bit more about these theories at the end of the chapter.

Naive Realism, and its related supporting theory Disjunctivism, is a currently expanding philosophical theory that has recently grown in popularity[79]. As an example of the things Naïve Realism claims I give the following three quotes, and it should be noted that the first quote is referred to by Kennedy (2007: 320) as a: 'fairly canonical statement of contemporary direct (or naïve) realism':

> 'According to naïve realism, the actual objects of perception, the external things such as trees, tables and rainbows, which one can perceive, and the properties which they can manifest to one when perceived, partly constitute one's conscious experience, and hence determine the phenomenal character of one's experience. This talk of constitution and determination should be taken literally; and a consequence of it is that one could not be having the very experience one has, *were the objects perceived not to exist, or were to lack the features they are perceived to have*. Furthermore, it is of the essence of such states of mind that they are partly constituted by such objects, and their phenomenal characters are determined by those objects and their qualities. So one could not have such a type of state of mind were one not perceiving such an object and correctly perceiving it to have the features in manifests itself as having.' Martin (1997: 83-84 - my emphasis)

and:

> 'This, I suggest, is how we should understand the naïve realist's claim that external objects and their properties shape the contours of the subject's conscious experience: they shape the contours of the subject's conscious experience by actually *being* the contours of the subject's conscious experience [and this is as opposed to just causing them more indirectly].' Fish (2009: 6)

and:

> 'The experience in a perceptual case reaches out to and involves the perceived external objects, not so the experience in other cases.' Snowdon (2005: 136-7)

I will now go through each quote in turn and say a few things about it. In the first quote we can readily see why I have adopted a terminology that suggests the object must play a

[79] C.f. Haddock and Macpherson (2008) and Gendler and Hawthorne (2006).

constitutive role in perception. And if this is taken literally as Martin suggests, this means we are literally aware of the object itself in some way. This immediately means this approach is quite obviously a direct realist theory, and I will say a little more about what this means in the next section. I have italicised a part of the first quote because this is a phrase that I will try to put pressure on in my Time-spread Argument to follow. Simply put, there seems to be many examples of perceived objects that no longer exist and yet we still perceive them, such as when looking at extinct stars (more on this in Section 5.3).

In the second quote we have another expression of the idea that it is the objects themselves that seem to be shaping our experience and not, for example, some indirect representation of them. In what follows we will briefly look at a move by Fish (2009) that defends Naïve Realism against the Argument from Hallucination by denying, against convention, that hallucination has phenomenal character. He also questions the idea that detailed phenomenal character can be locally supervenient. If this is plausible then it threatens to be eliminativist about, not only the phenomenal character of hallucination, but also that of phenomenal imagination. And hence may represent a potential threat to my assumption that there is such a thing as phenomenally experienced imagination in the first place. I will explain why this seemingly unintuitive step is motivated at the end of this section and give one example, based on after-images, of why I think it isn't necessary to take up this extreme position in order to defend Naïve Realism.

In the third quote we have an explicit expression of the 'reaching out' terminology, where the mind is said to reach out and put one directly in contact with external objects (and see the opening quote from Bonjour 2004). This also illustrates one way in which Naïve Realism can clearly be seen as a Disjunctivist Position, because it seems to be saying that this reaching out is unique to perception and is not the case in other experiential states, such as perhaps hallucination and phenomenal imagination. Hence we can initially see why vision and other vision-like states might be classed as very different kind of things with little or no common overlap. Martin (2004: 38) suggests that: ‘The prime reason for endorsing Disjunctivism is to block the rejection of a view of perception I’ll label *Naïve Realism’*, and as far as I can tell all modern naïve realists are also Disjunctivists. So having explained the quotes above in a bit more detail, perhaps we can now briefly look at the main claims of Disjunctivism.

Specifically Disjunctivism is employed to defend against the Argument from Hallucination[80] where it now becomes possible to deny the following: when two mental states are phenomenally indistinguishable, this means that they must be the same kind of thing. Analogously a lemon fruit and lemon shaped bar of soap, may be superficially indistinguishable on sight, but it would be a mistake to read anything more into this in terms of deeper ontological categorizations. Fish (2005: 122) refers to this as the ‘Decisiveness Principle’, which is where we infer from phenomenal indistinguishability that there must be some further commonality between the items in question. He suggests that: ‘…endorsing the decisiveness principle leads straight to a non-disjunctivist theory’ and ‘that there is no *contradiction* in supposing that we might be unable to tell the difference between two things which are nevertheless fundamentally *different*.’ Hence, agreeing with Fish, we must be wary of arguing for a more Conjunctivist Position based simply on these superficially matching appearances. They may turn out to be the same thing, but both sides can agree that a more detailed argument is required to establish this.

[80] It is also used to defend against the Argument from Illusion. However this aspect is much less clearly developed in the literature and even the position of illusion either side of the disjunct is also debated between researchers. Hence I will stick to discussing hallucination in what follows: but c.f. Byrne and Logue (2008) and Thau (2004) for a good discussion of this. And note I follow Thau in treating illusion as a kind of perception, all be it one that has gone wrong in some way: analogously Thau (2004: 248) suggests a mis-kick of a football is still a ‘kick’, rather than something else, like a complete miss.

So the main point of appealing to Disjuntivism as I see it, is to preserve a special place for veridical perception as a thing that does put us in direct contact with the external world. This is contrasted with, for example, a vivid global hallucination that would not do this, even though the two might be phenomenally indistinguishable. To me, this suggests that while veridical perception may only be compatible with Selective Approach in this way, a naïve realist *might* agree that a phenomenal hallucination could be characterised under a more Generative Approach. At least I think taking this route doesn't obviously seem to violate the main concerns of Naïve Realism or Disjunctivism and has certain other benefits to be discussed next. In support of this claim, consider the following quote:

> "…[Disjunctivism] is not inconsistent with the view that there are some experiences among the non-veridical ones which fit the characterisations offered by sense datum or intentional theories." Martin (2004: 52)

I take this as at least accepting the possibility that other vision-like states could fall under some theories with more sympathy towards a Generative Approach. This sympathy seems to be reasonably explicitly stated in the above, but the Disjunctivist literature in general focuses more explicitly on hallucination because of it was specifically developed as a defence against the Argument from Hallucination. However we might now also want to ask, particularly within the context of this thesis, how this dialectic might comment on phenomenal imagination as a simulation process?

My interpretation is that phenomenal imagination would fairly obviously fall on the non-veridical side of the disjunct along with hallucination, and hence could also be treated under a more Generative Approach. Consider the following argument in support of this: if the original reason for explicitly stating this disjunct is in terms of hallucination, and this is because it is a threat because it could possibly be mistaken for a veridical perception; then I see no reason why anyone would deny this disjunct in terms of phenomenal imagination, given that it is even less likely to be claimed to be the same thing as a veridical perception. In other words, since phenomenal imagery is not even phenomenally indistinguishable from veridical perceptions, there seems even less prima facia reasons for claiming they might be the same thing. If that thought process is sound then it seems reasonable to me to assume that the Naive Realist *could* accept that phenomenal imagination is a different kind of thing from veridical perception, and be fairly neutral as to whether this falls under a Generative Approach or not. So as long as phenomenal imagery remains a different kind of thing to a veridical perception and doesn't threaten its status as giving direct contact with external reality, then naïve realists could plausibly remain neutral to how it is developed further.

It should be noted that I have qualified the statements above with 'could' and 'might', because generally the Disjunctivist debate only explicitly refers to hallucination. And apart from the explicit quote regarding the difference between perception and imagery given in Section 1.4, from Martin (2001: 273-274), this is merely my interpretation of the situation on behalf of other naïve realists. Hence it merely represents one possible way of approaching the disjunct between perception and other non-veridical visual-like states like phenomenal hallucinations and imagination. However given the reasoning process above I think this is a fairly plausible route to take and may at least act as a good starting point for extending this debate towards the imagination end of the vision-like experience spectrum. Therefore I feel reasonably justified in giving the following illustration in Figure 5.1 of *one possible way* of describing visual imagery within a Selective Approach to perception. It should be noted that this actually involves proposing a mixed approach, where we adopt a Selective Approach for perception and Generative Approach for phenomenal hallucination and imagery:

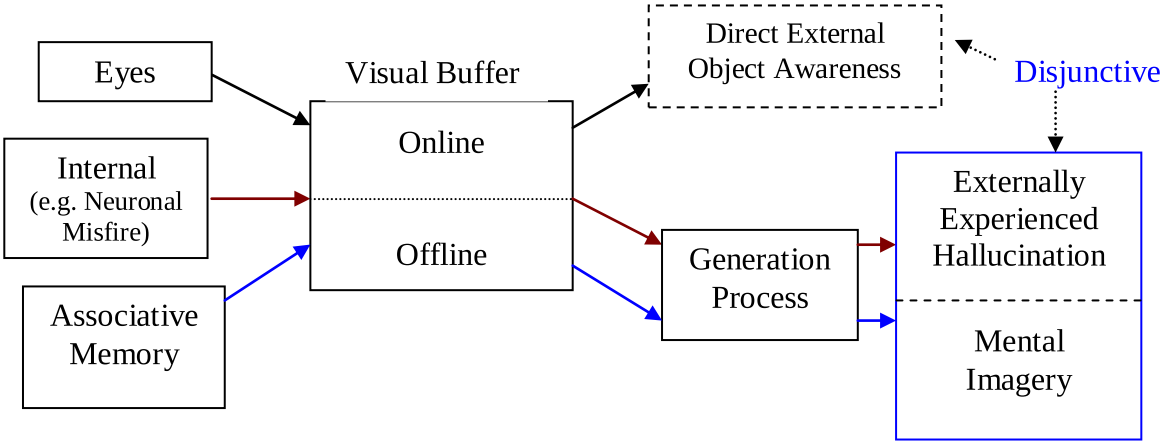


Figure 5.1 – Diagram illustrating one possible way that visual imagery could be described within a Selective Approach to perception

In the diagram above I have tried to illustrate a possible way of explaining the disjunct between veridical perception and other vision-like states, by very obviously placing them in separate output streams. This has the benefit of fitting with the compelling overlapping mechanism evidence on the input side, which has been discussed in the first part of the thesis. In other words they can both utilise the visual buffer in common, as long as how they are treated afterwards constitutes a big enough ontological difference to justify the disjunct. The main difference in the output stream that I think could justify this, is the way this information is treated once it leaves the visual buffer. Specifically, in the visual case this information acts to somehow put you in touch directly with external reality, but in visual imagery and hallucination case, the brain generates the phenomenal character itself. In this latter generative stream there is a further difference, in that the internally caused (potentially due to neuronal misfiring) hallucinatory information is somehow treated as being online and hence experienced externally; or at least is (mis)taken as having occurred externally. And this is opposed to the visual imagery stream, which is caused internally as previously discussed (e.g. by an act of will) and goes through the generation process but is not then experienced externally. And while I appreciate there are a lot of details I haven't mentioned in this diagram, I do not want to go into more detail here[81].

I do not intend to spend more time on explaining Figure 5.1, as discussed earlier, I do not favour this kind of approach in the first place, and merely offer this description here as one way that this could be developed if so desired by a defender of a Selective Account. Saying that I do think it offers one fairly reasonable way of explaining how the main claims of Naïve Realism can be preserved in terms of direct veridical perception, whilst still allowing the possibility that other vision-like phenomenal states could be locally supervenient and generated somehow by the brain. And as mentioned earlier this seems at first sight to remain compatible with the known empirical evidence from visual science as discussed in the last two chapters. So there are some benefits to adopting this kind of approach for the naïve realist and perhaps therefore offers one potential platform from which it could be developed further. This is particularly pertinent since Disjunctivism is sometimes accused of not offering much in the way of explanation of the non-veridical side of the disjunct[82].

Having said that, one interesting implication of this approach is that it perhaps indicates that phenomenally experienced hallucinations cannot be produced in a proximally causally matching way to veridical perceptions: e.g. by the exact same overall brain state. This is because of the extra generation step involved in hallucination that is not present in perception, which I would suggest requires at least some minimal change in the underlying

[81] For example, I have placed the hallucination stream arrows as falling in the cusp of the online and offline streams to indicate that this would need to be discussed further in order to refine this placement. I also haven't discussed how this process would be instigated in the first place and have just suggested it may be due to an 'internal neuronal misfire' of some kind: perhaps due to an in-balanced neuro-chemistry, which in turn could be due to, for example, tiredness, starvation or a drug intake.

[82] See Johnston (2004), Dunn (2008) and Sturgeon (1998) for expressions of this sentiment. But note this is becoming less acute as Disjunctivism gets more attention in the literature. And also note that this mixed kind of approach may be implicated in Noe (2006) (see also Prinz 2006).

neurology. This could possibly be seen as a benefit of this approach, because accepting that these two phenomenally indistinguishable yet fundamentally different mental states can be caused by an identical underlying neurology has lead, in my opinion, to some strange consequences in the literature.

For example, I think that accepting phenomenal indiscriminability and a proximally matching causal base, leaves very little room to explain where the ontological difference between perception and hallucination comes in. So much so that in Fish's (2008, 2009) case, this has lead to the unintuitive denial that hallucinations have any phenomenal character at all. This is suggested because, if we accept both the above constraints, this might indicate that the phenomenal character involved in hallucinations could 'screen you off' from the external world in the perceptual case[83]. So to deal with this, Fish suggests that perhaps in hallucination it is just that you merely *believe* you are experiencing something and you are in fact fooled into believing it has phenomenal character and therefore it does have some 'felt reality' associated with it. I referred to this as a doxastic hallucination and Fish spends some time explaining how this kind of thing could fill the same functional role that a more traditional phenomenal hallucination does. He also spends some time challenging the assumption that the brain can support detailed vision-like phenomenal character and hence denies a certain form of the local supervenience principle. This eliminativism of the phenomenology of hallucination it seems to me could perhaps also lead to denying that of vivid visual imagery[84], and I think this also a very unintuitive claim since I personally experience quite vivid phenomenal imagery; as I assume many other people think they do. So this leads to a situation where Fish is defending the fairly intuitive

[83] See Martin (2006), Fish (2009), Johnston (2004) and Dunn's (2008) reply to Johnston for more on the 'screening off' problem. Especially useful is Johnston's 'Surgery Example', which illustrates this point well.
[84] I have had confirmation of this potential threat to the phenomenal character of imagination in personal communication with William Fish.

position of Naïve Realism (see later), by adopting a fairly unintuitive position that denies the phenomenology of hallucination and imagery.

However in my opinion, perhaps the biggest problem for Fish's approach is that it seems to also deny the phenomenology of other experiential states like phosphenes[85] and dreams, and even struggles to explain after-images. All of which are generally agreed to have quite detailed and vivid phenomenology, although they may differ form visual states in other ways: e.g. they fade quickly or are temporally disjointed. Confirmation of this stance is expressed in the quote below:

> 'I see no reason why even phosphene experiences, after image experiences and the like, should not submit to the same kind of explanation [as hallucination].' Fish (2009: 134 – footnote 12)

The problem with this latter claim is that although we cannot just summon up a hallucination to check our phenomenology, we *can do so* with an after-image. To see how unintuitive this claim is first hand, follow the instructions below that lead to what I would suggest is the generation of a fairly vivid, complete and potentially indistinguishable after-image.

Here you need to look at the left hand white dot for a while, until you have the sense that there is a complete red square present in the figure on the left (it may help to cover the right hand side of the image with your hand). Then you look at the orange dot on the right and observe the left side phenomenology and how closely it matches your perception of the right hand side of the orange dot. And note that this may involve shifting your attention while keeping your eyes focused on the orange dot:

[85] Phosphenes are the coloured experiences one has when you close your eyes and put light pressure on the sides of your eyes for a minute or so.

Figure 5.2 – After-image example designed to test how well you can discriminate it from a closely matching perception

What you should see on the left of the orange dot, is a quickly fading version of the image on the right hand side of the same dot: i.e. a green square over four black circles on a grey background. And while it is true that these two experiences do not exactly match, they are fairly similar in at least general colour and shape[86]. It seems then that the implication of Fish's position is that: while it would accept that you are having the phenomenal experience on the right of the orange dot (as a veridical perception); it would deny the phenomenology on the left of the orange dot to do with the after-image (in a similar way to a doxastic hallucination). That is, on the right you are in a normal phenomenal experiential state, but on the left you just believe you are, but are actually mistaken about the apparent phenomenal character you seem to experience. This is even though they could theoretically be made to appear completely indiscernible (and see footnote 86). This seems highly implausible to me and leads to several other problems that I can't go into in detail here[87]. Suffice to say that if the mixed Selective and Generative Approach to vision and visual imagery that I have proposed above (c.f. Figure 5.1), precludes the possibility that

[86] They could perhaps be made to match more closely by someone with better graphic design skills than mine and even made to fade at the same rate on a computer screen. And note I have found this after-image much harder to generate from a paper print out, as compared to a version presented on a computer screen. So I refer the reader to the original article this came from which can then be viewed on a monitor as a PDF file (see Shimojo et al 2001- and note I have adapted their example slightly to illustrate my points here).

[87] But see my book review of Fish (2009) for more on this issue (King forthcoming). As a brief example of the concerns expressed there, it seems that accepting this interpretation might make us worry that we could also doubt the phenomenal character of the veridical perception. If this is so it seems possible that we are actually philosophical zombies and just believe we have phenomenal character in all cases!

perceptions and hallucinations can share a common underlying neurology, and hence avoids needing to deny the phenomenal character of hallucinations and after-images, then this may not be such a bad thing.[88]

The other good thing about using this particular after-image is that it also illustrates a phenomenon known as colour bleeding, where non-coloured areas of the original seem to get filled in by the mind: e.g. to form the completed square shapes. This occurs, not only in the original image on the left, but also in the after-image of Figure 5.2. This could be interpreted as an example of where the mind is anticipating what it should see and 'filling it in' phenomenally. It is tempting to link this to the discussion of visual emulators and the internal filling-in of the visual buffer discussed in Section 4.3.2. But attempting this cross level correlation is a lengthy, complicated and possibly a controversial process[89]. Suffice to say that the explanation just mentioned, involving a correlated filling-in at the phenomenal and sub-personal level, is certainly one possible interpretation. However the main point of introducing this example is that it seems to offer some support to the claim that the brain generates vision-like phenomenal character in at least some cases. Hence this may further vindicate adopting the mixed Selective and Generative Approach as proposed above.

So in summary, the mixed approach illustrated above has the benefit of fitting with the known empirical data, at least on the input side, as described by my defence of the Kosslyn Model previously. It also has the advantage that it makes the ontological difference

---

[88] For the record, Fish (personal communication) agrees that he could fall back on something like the mixed approach discussed above. However he thinks this is a slightly weaker position and prefers stay with the more radical option that he has argued for already. Therefore depending how things go for Fish's theory, this option may become more attractive in the future, but it is hard to tell at present. For the record I don't really see why a Disjunctivist need accept that perception and hallucination can occur in a causally matching way. It is certainly not a constraint that is insisted on by any opponent, but one that they seem to accept voluntarily as far as I can tell (c.f. references in footnote 83).

[89] See Noe et al (1997) for a review of the range of interpretations of this phenomenon. I think my interpretation is not ruled out by them but it is certainly cautioned against, in much the same spirit as Hurley's (2008) 'interlevel isomorphism' warning that I discussed in Section 2.1. I will also discuss this further in Chapter 7.

between veridical perception and other vision-like states very apparent and can therefore obviously justify their treatment under a Disjunctivist Position. Unfortunately I do not have the space or motivation to develop and explain this formulation further, because in the next two sections I will give my reasons for rejecting Selective Approaches like Naïve Realism in the first place. So the above is just offered as one potential way to develop a theory of visual imagery as an offline simulation process, given a Selective Approach to perception. If my concerns about the Time-spread Argument discussed below could be ameliorated, then this is possibly something I would enjoy developing further in the future. But until then my preferred avenue for developing a full theory of STIm, involves the adoption of a Generative Approach. Defending this move is what we can turn to now by, first attempting to neutralise some of the main motivations that I think there are for adopting Naïve Realism in the first place (Section 5.2), and then rejecting it outright based on the Time-spread Argument (Section 5.3).

5.2 – Neutralising some general arguments for Naïve Realism

Before looking at the Time-spread argument against Naïve Realism in detail in the next section, I think it is worth explaining why I take that argument to be the most crucial one to discuss in this context. This is because I feel many of the other arguments for and against Naïve Realism are less compelling ones, or have been adequately dealt with for now. For instance, it seems that the appeal to Disjunctivism offers some immunity from the Argument from Hallucination, or at least this is a strategy that is currently under debate and is garnering much support. In what follows I will briefly explain why I don't feel motivated to defend Naïve Realism on other grounds and this will also act to help justify why I am happy to adopt a much less intuitive indirect realist and Generative Approach in what follows. So for now we can briefly look at why someone would willingly adopt a theory that has the word 'naïve' as part of its name, and I think these fall into two main categories to do with its apparent *intuitive* and *epistemological* benefits[90]. I will discuss each of these in turn below.

The **first** main motivation that I think there is for adopting Naïve Realism is that it is the most intuitive position to adopt. It just patently appears to us that we are directly in contact with external objects and it is these objects that form a constitutive part of our perceptual experience. This can be perhaps best be expressed as two separate but related intuitions. Firstly there is the '*Openness Intuition'*, which claims that it just seems like our perceptual states are 'open to the world' and it is the world itself we are aware of (c.f. Crane 2005). The second is the *'Transparency Intuition'*, which claims that in most normal perceptual experiences, it doesn't seem like we are observing the world through any intermediary states. If it turns out we are, then these are usually transparent to us and it's as if we see straight through them; like a really clean window. The simplest explanation for both these intuitions is that we are actually just directly aware of the external world and that in fact there is nothing transparent through which we see. Hence transparency may actually be a misnomer in this situation[91] and the natural conclusion to reach from this is that both these intuitions favour Naïve Realism. Consider this quote from Kennedy emphasising this point:

> 'Representationalists take transparency to support their theory and to work against the qualia theory. In this paper I argue that representationalist assessment of the philosophical importance of transparency is incorrect. The true beneficiary of transparency is another theory, naïve realism. Transparency militates against qualia and the representationalist theory of experience.' Kennedy (2009: abstract)

[90] And see Pautz (2007: 504) for a similar division of the field into intuitive vs. epistemological arguments used in defence of Naïve Realism. This is not to say that there are not others but I interpret them as the most important ones.

[91] I take this point from Kennedy (personal communication), because perhaps for Naive Realism there is nothing literally 'to see through' or to 'be transparent' as such. The above represents my use of the term 'transparency', but see Siewert (2004), Kind (2003), Loar (2003) and Crane (2006) for how this word has been variously used and abused.

In general agreement with Kennedy I do not intend to dispute that our basic intuitions favour the naïve realist and not the Representationalist; this point is readily conceded. However I would point out that this is mainly because I don't think it concedes much of importance. Consider the strength of the intuition that the sun revolves around the Earth: one can concede that this is the naive intuitive position, while still holding relatively sophisticated scientifically based reasons for denying it. And further, even after you are aware of these facts, it is still very hard to see this situation as the Earth rotating relative to a static sun and hence it remains difficult to change your intuitions. Analogously our intuitions about perception will probably always favour Naïve Realism, even if it turns out we can eventually prove that it is false, which therefore devalues it as evidence for or against a position in the first place. Saying that, I accept that ones intuitions should only be doubted if there are fairly strong reasons to do so. However *I do think* that there is already enough evidence to seriously doubt whether our intuitive take of perception should be trusted and much of this is also science based and is why many researchers in this field are not naïve realists[92]. Hence I suggest people should be wary of leaning on the Openness and Transparency Intuitions too heavily and much of the next section is aimed at explaining why I think this reliance is a mistake and to outline reasons for resisting these intuitions. This will hopefully have the effect of removing one of the main motivations for defending a naïve realist position.

Related to all that, I think the **second** main motivation for defending Naïve Realism, is that it seems to give us a more assured and direct epistemological access to the external world. The idea being that if external objects are mainly as we perceive them and it is the objects themselves that we directly perceive, then this acts as a relatively solid basis for our knowledge claims about reality. This is relatively solid when compared to, for example,

[92] C.f. a collection of papers gathered together under the 'qualia' banner in Wright (2008), many of which appeal to scientific reasons for rejecting direct realism. And see Chapters 6&7 for more on this.

indirect realist theories where we may be mistaken about a large part of our experience and perhaps actually truthfully know very little about external reality: e.g. if a Generative Approach is true then we seem to be only aware of states generated by our brain and are therefore cut off in some way from the external world in itself[93]. This point is related to the usual 'veil of perception' based arguments used against indirect realist theories. I refer to this motivation for arguing for a naïve realist position as the *Epistemological Desire Argument*. The quote below may be an example of this phenomenon and is also an example of Johnson's use of his 'conjunctive' terminology:

> 'Despite the venerable credentials of the argument, we should reject the Conjunctive Analysis of veridical sensory awareness… What is odd about the analysis is that it entails that the objects of hallucination are present to us in a way that external particulars cannot be, *even when we are seeing external particulars*. We should instead hold out for a view to the effect that when we see, or more generally, sense, external particulars those particulars are no less "directly" present to us than anything is in hallucination.' Johnston (2004: 119)

The idea that we should 'hold out' for a perceptual theory that fits our epistemological desires, seems to risk precluding some very plausible theories, simply because you don't like their potential consequences. And while I generally sympathise with these epistemological desires (i.e. I would also *like* direct realism to be true), I don't think this is a valid form of argument because it is at risk of putting the cart before the horse. What we desire, whilst possibly being a valid initial motivational factor, is not actually an argument in defence of that same theory; that is any more than my desire that the Earth be flat is an argument for flat-Earthism. The moral being that we must discount what we would like perception to end up delivering epistemologically, from influencing which theory we

---

[93] This is why Martin (2002: 421) calls all indirect realist theories 'error theories'. But note that believing the Earth revolves around the sun is also an error theory. So this must just be a descriptive rather than a pejorative use of the word 'error'. It may well be that a kind of error theory of perception is in fact the correct one and this is something I try to defend here.

initially favour. In that case, in my opinion, epistemological desires must be confined to a motivating factor to defend the theory, not a defence of the theory itself.

Conversely accusations of 'putting up a veil of perception' must be discounted as arguments against any indirect realist or Conjunctivist theory, simply because this may be an unavoidable consequence of this theory if it is indeed correct. It may not be, but we shouldn't resist concluding a theory is correct simply because we don't like its implications. This is why I suggest we should not concern ourselves with the epistemological implications of the theory too early, because I suspect that at times this leads to a discounting of the scientific evidence and possibly leads to a misleading spin on the metaphysical arguments. As Rey (1998: 455) points out: '…surely one must be wary of such attempts to buy epistemology by selling out metaphysics'.

Added to this is the fact that even if indirect realism were true and we don't have as good an epistemological access to the world as we would like, it still seems like we have just what is necessary for survival evolutionarily speaking. That is if indirect realism is true then it seems that this is all we need for survival, and this is better than no access at all and hence there is no need for absolute direct access. So perhaps to desire more than the basics needed for survival might seem anachronistic or anthropocentric and hence I don't really feel the impact of veil of perception related objections. We seem to 'know' just enough to survive and thrive as a species and hence desiring any higher kind of 'Knowledge' may just be a pre-Darwinian remnant of our outlook on the world and our place within it. And see also the opening quote from Bonjour (2004: 363) for a statement of how limited the actual epistemological benefits of Naïve Realism may be given an internalist epistemological approach. To paraphrase what I think Bonjour is saying there: it seems that even if Naïve Realism is true, you still can't tell when you are hallucinating or perceiving, and therefore would be unable to tell what is a veridical perception or not. And hence our whole perceptual experience could still be in global error like that of a BIV. Hence given an internalist epistemological approach we are still no better of at combating global scepticism if we adopt a direct realist position[94].

Related to the discussion above, since Naïve Realism involves such a simply construed close contact with the world, I accept that it has the strongest claim to being called a direct realist theory. Indeed I greatly respect its clarity and the candour of its supporters on at least this issue. But perhaps this means that in comparison any other theory that does not claim this immediate contact with reality, must at least be seen as a 'less-direct' theory. And even though less direct theories may seem to be less intuitive, just being labelled as an indirect theory is not necessarily a bad thing as long as its well motivated. Consider the following from Fish on this point:

> '…So calling a theory "indirect" should not, I maintain, carry with it any negative connotations which may have been associated with it for these reasons [such as the veil of perception]. And it may well be, therefore, that some version of a non-disjunctive theory such as the representational theory turns out to be the best theory of perception we have. All this paper aims to show is that, if that were the case, our best theory of perception would fall into a category which is justifiably, but non-pejoratively, known as "indirect realism".' Fish (2004: Section 4)

In what follows I will therefore defend an indirect realist theory on the understanding that, while this may seem unintuitive, this does not a necessarily immediately make it a weaker position to adopt, and there may be compelling reasons for accepting this categorisation. I also think that accepting the indirect tag in a non-pejorative way really helps be clear up

---

[94] Although given the above, I can appreciate how Naïve Realism might improve our epistemological access if a form of Contextualism is adopted (c.f. Sosa 2000, DeRose 1999), and we confine our discussions to a more local form of scepticism. This could take place under the assumption that we are not a BIV and could then look at epistemological concerns within this restricted context. So I am aware that this issue is a complex one, but I cannot really allocate it more space in this thesis: and see the comments that come next in the main passage.

the terminology in this area, where many different theories seem determined to claim directness, even when it is not obvious how this is supported. For instance, I think this area has perhaps been muddied by theories that refer to themselves as 'Representationalism', and I join the naïve realists in being sceptical of how they can back up these claims of directness [95]. This will be discussed further in Chapter 6.

This completes my attempt to neutralise the two main motivations for adopting Naïve Realism as I see them. Before moving on it should be noted that the two discussions above are only aimed at diluting or neutralizing these motivations and are not presented as knock down arguments against them. What I hope to have explained is merely why I do not feel the same draw towards this kind of theory and this is partly why I am happy to adopt a Generative Approach. So the arguments above are merely designed to give the reader an initial indication of why I think I can neutralize these motivations and why other issues, like the Time-spread Argument discussed below, tip the balance for me towards Generative Accounts. So now that we have introduced Naïve Realism and I have explained its main motivations and why I resist them, lets look at what my main argument against it is.

5.3 – The Time-spread Argument against Naive Realism

In this section I will take the classic Time-lag Argument normally cast against direct realist theories and make it more acute by spreading and sharpening it up. The eventual aim is to show why I think these kind of theories cannot deal with this problem whilst still remaining physicalist or without introducing implausible new physical properties; both of which *I do take* to be pejorative claims. Specifically I will show how the Time-lag

[95] For example Tye (1995) and Dretske (1995) and other theories within the school now called 'Representationalism'.

Argument can be made more acute for this kind of theory by introducing the idea that we are actually aware of objects as they were spread backwards through time. I refer to this new version of the Time-lag Argument as the Time-spread Argument, and I will focus on applying it against modern Naïve Realism, as I see it as the most unambiguously direct realist theory on the market. I will particularly put pressure on the naïve realist claim that external objects must literally constitute part of your experience and hence just *are* what make up the contours of your visual states (c.f. the quotes at the start of the chapter). This is often described as 'your mind reaching out to put you in direct contact with objects' and I will try to show that this is a highly implausible metaphysical claim. This also puts pressure on the idea that our perceptual states widely supervene to include the object, because the properties of the object may have drastically changed by the time you perceive them. The most extreme case of this is where the object may have ceased to exist entirely many years before you perceive it, as in the classic time-lag counter example of seeing an extinct star.

I will take Le Morvan's (2004) position as my representative direct realist defence against the Time-lag Argument. His paper draws together eight main arguments against direct realism and gives a standard defence to each of them based on a review of the literature. Hence he presents these defences as the united reply from direct realists generally and I think therefore it appropriately characterises the best modern exposition of a direct realist defence against the Time-lag Argument[96]. He summarises this defence as follows:

> 'It's important to keep in mind here that Direct Realists need not be committed to the claim that we can *now* be aware of the no-longer existent object as it is *now*, but only

[96] This is certainly in line with what I have gathered in personal communication with modern naïve realists. And see Robinson (1994: 80) for a good introduction to the classic formulation of the Time-lag Argument. Note that I am aware of another recent relational direct realist attempt to reply to the time-lag argument given by Langsam (1997). However I think Djukic & Popescu (2003) have dealt with this adequately and my arguments here also deal with it indirectly and create new problems for Langsam.

that we can *now* be aware of the once-existent object *as it used to be.*' Le Morvan (2004: 224)

The first thing we can note from this defence is that it immediately weakens any appeals to the naïve intuitions we have about perception. This is because my naïve intuition is that I just see objects as they are now, and not that I see them as they were in the past. However this is only a mild weakening, a much stronger critique is that being aware of things as they used to be in the past is by no means a trivial metaphysical claim, and we will see that even this refined defence is still possibly inaccurate and implausible. So in order to develop this stronger critique further, I will introduce the Time-spread Argument by way of developing the 'Blipper Example' in the next two sub-sections and use it to help illustrates my points. I will then look at how Le Morvan's defence can be interpreted under different theories of time in sub-section 5.3.3 and show how this defence seems implausible irrespective of your position on the nature of time.

5.3.1 - The Blipper Example and the Time-spread Argument

Lets begin by noting the different light travel times from certain visible objects as a way of illustrating the spread in time-lag between objects at varying distances from you. So imagine it's twilight and you are enjoying the last bits of the sunset across a panoramic view of distant mountains, but you can also see the Moon and a few early stars, some of which are planets, all at the same time. I appreciate this is a fairly unusual situation but this does not affect the main points I want to make here. Some examples of the travel times of light reflected or emitted from surfaces of these various objects to your eyes are as follows:

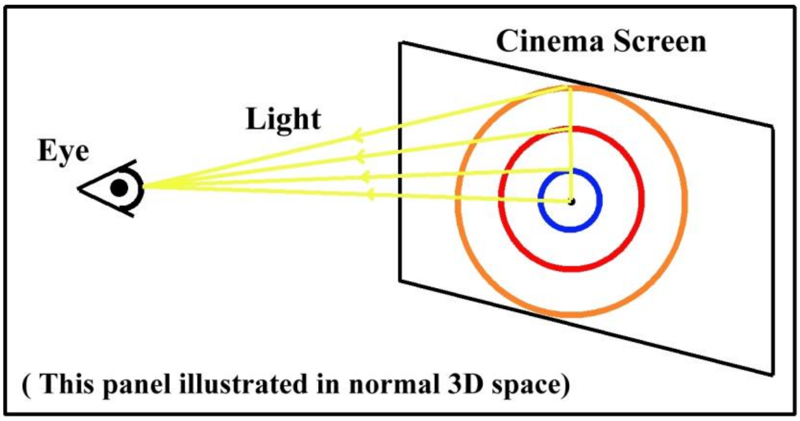
Object	Light Travel Time (approximates)
Stars	millions of years
Planets	e.g. 4mins (Mars minimum) and 32mins (Jupiter minimum)
Sun	8 mins
Moon	1.3s (assumed 1 second for ease of reference below)
Anything on Earth	0.01s or less (one hundredth of a second)
N.B. 0.1s (one tenth of a second) = Assumed minimum time interval discriminable by the human eye (this is approximate – see below); and speed of light = 300,000km/s	

Table 5.1 – Table of light travel times from various objects to your perspective given the situation described in the text where they are all visible to you at once at twilight

In the table above the light travel time range is between, a few millions of years, to less than one hundredth of a second. This means the view you are enjoying in the example above is actually a conglomerate of objects as they were at different times in the past. I think this immediately counts against the Openness Intuition introduced above, because you are not aware of the world (and universe) in a formation that it has ever been in. Certain of the stars you view have not even existed at the same time as each other, so this particular configuration you are experiencing is unique to you where the light is converging on your eyes and not anything that has ever existed in the external world. Hence it seems implausible to me to suggest you are perceptions are ‘open’ to the world in the way this intuition initially suggests. This phenomena is what I call *inter*-object time-spread, but there also exists a more subtle form I call *intra*-object time-spread.

To illustrate intra-object time-spread, imagine looking at a large blank cinema screen. Lets say you are seated exactly opposite the middle of the screen, so that the shortest distance from you to the screen is a straight line that comes out at right angles from the centre of the screen to a point between your eyes. Now imagine a small circle around the centre point that slowly grows bigger until it goes off the edge of the screen all together[97]. Now as the circle increases in size, the distance of any point on its circumference, to where you are sitting, is also increasing. This is because it is along a hypotenuse of an imaginary right-angled triangle that goes from any point on the circle to the centre of the screen and then out to you and back to the circle. Hence the travel times of light from the circle to you are also very slightly increasing as the circle moves outwards. If you represented this on a time axis then it would look like a cone spreading out backwards through time, like this:

[97] Note this is also another good example of how visual imagination can help you understand what I am talking about. And there is another example of this in two paragraphs time.



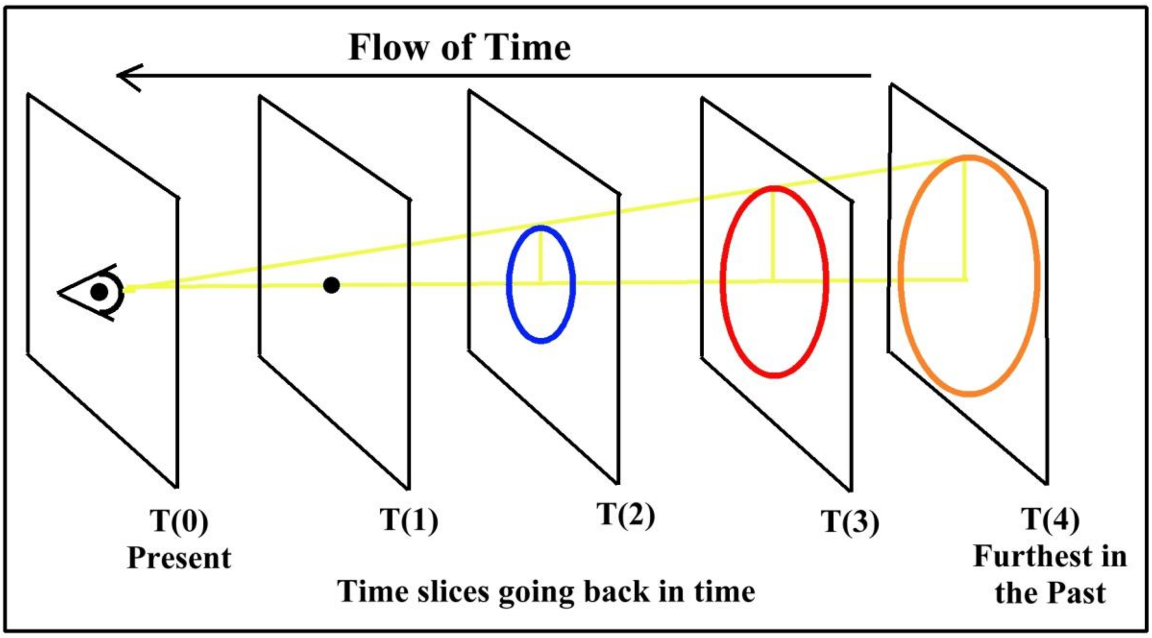


Figure 5.3 – Illustration of the cinema screen Time-spread example, where: T(1), or the small black circle, represents the point closest to you on the screen and hence correlates with the fastest light travel time; and T(4), or the large orange circle, represents a point farther away from you on the screen and hence correlates with a larger travel time; and T(2), the blue circle, and T(3), the red circle, are intermediary stages.

The diagram above generalizes to any object that has any facing surface area relative to your perspective, because there will be a slight differential in travel time from different parts of the object that are different distances away from you. Since most objects we can experience do have surface area, we can safely say that in nearly all cases of perception there is a fractional amount of intra-object-time-spread. But how big is this fraction?

Well, in the table above I have indicated that, for the sake of argument, I will take the smallest discernable temporal interval in visual perception to be about 0.1s (or one tenth of second). To get a feel for this imagine a light flashing ten times a second and trying to count the flashes. I think just from this imaginary exercise we can agree that the individual flashes are close to being indiscernible at ten per second. To illustrate this further we can probably agree that twice a second would be easily discernable, and twenty times a second they would be definitely indiscernible. It is worth noting that it is surprisingly hard to get an accurate figure on this minimum threshold because it seems to vary between individuals and with the intensity of the light source. Measuring this minimum also doesn't seem to be anything that anyone has spent any time researching specifically and hence this is just a figure I have gathered from various online sources and extrapolating from phenomenal experience. Fortunately the exact figure doesn't matter too much because anything you could possibly observe on Earth has a travel time that is way below this threshold. For example, in the cinema screen example, the maximum travel times are many factors of ten below this 0.1s threshold. The point is that you wouldn't notice this effect from any earth-bound object even if you tried to. So to develop these ideas further we can introduce a useful object called a 'Blipper' so that we can move it away to greater distances that are not earth-bound.

The Blipper Example

A Blipper is a variable intensity light that flashes a strong blue light and looks significantly different when it is not turned on (e.g. bright blue vs. see through). The Blipper is set up so that it can work out how far away it is from you because it is wirelessly connected to a GPS receiver mounted on your head. This distance-to-your-eyes measurement is then used to change the frequency that the Blipper flashes so that its duration is exactly the same as the time it takes for light to travel to you. The net effect is that when you perceive the Blipper as being on, it is actually always off and vice versa. In other words, as soon as the light from it reaches your eyes, it turns itself off, and as soon as the light stops arriving at your eyes from the last flash, it turns on.

> N.B. The Blipper example can also be extended to be a situation where its actual shape also changes. Or more exotically it gets teleported in front of you and then disappears at the required frequency. This is introduced to pre-empt defences against this example that might concede the point in terms of colour, but deny it in terms of shape. In other words this example can be adapted so that both the shape and colour you should perceive are not present in the external world at the time of perception. And note it is not the actual possibility of this set up that is at question, but rather how your perceptual theory would deal with it if it were in fact possible.

Now obviously at earth-bound ranges these changes will not be noticeable because the Blipper will be flashing at a frequency way below your discernable threshold. It may be that it looks a bit lighter than an equivalent permanently-on-Blipper next to it, because the light intensity is halved, but this is not a feature I need to appeal to here to defend my position.

The point of introducing the Blipper is that we can now move it gradually away from you, while at the same time gradually increasing its light intensity (and possibly size). This is so that it appears roughly the same to you and you can still see it when it is at a much greater distance, such as on the Moon. At this point we'll assume that the travel time to the Moon is about one second and therefore the Blipper will be flashing on and off at about once per second; which is now well above your discernable threshold of 0.1s. It is here that I think the naïve realist has problems with an explanation. Because all the time you are perceiving bright blue light there is in fact none in the external world and when you don't see anything but the Moon (because the Blipper is not visible at this distance when turned off), there is in fact a bright blue light shining at you.

The problem for the naïve realist here is that there is nothing in your present environment that is instantiating the kind of properties that match up with your perception at the time, and hence nothing that seems to be able to literally constitute your experience in the appropriate way. To put it another way, the features that you are experiencing are not currently present in the external world and hence unavailable to form the *actual* contours of your perception. So we must literally be directly perceiving the Blipper and the Moon as it was 1 second ago. That defence is fine as it stands so far, but I think there are two places I can now put further pressure on this line of argument. The first is to suggest that we are *not even aware* of things as they were in the past due to intra-object-time-spread, and the second is that seeing things as they were in the past is a controversial claim; and I will deal with each of these in turn in the next two sub-sections.

Before doing that I will need to make one assumption on behalf of the Naive Realist, and this is to assume that they are not going to suggest there is a different process underpinning perceptions of earth-bound objects compared with those at distances away where the time-lag is theoretically discernable. I say this because I want to exclude the possibility that these two versions of a perceptual experience will be treated or classed differently. It seems to me that exactly the same perceptual relations and mechanisms are present in these two cases (i.e. the Blipper on Earth or on the Moon) and if the earth-bound ones reach out to put you in direct touch with external objects then I would suggest it makes sense to assume that the non-earth-bound ones do too.

For instance, I would be sceptical of a move to treat the non-earth-bound examples as illusions, since there is a continuous progression as the Blipper moves away from you to the Moon[98]. It also seems that there is no deviant travel paths to appeal to (like the stick in water) and no cognitive interpretation effect to appeal to (like the Muller Lyer Illusion)[99].

[98] This move has been suggested to me by one naïve realist as a possible defence and specifically rejected by another because these non-earth-bound perceptions can still be defended under a Le Morvan's style defence. So it is debatable if this option would be favoured eventually; I merely offer the above as a pre-emptive reply if it was favoured.

[99] I refer to a useful division in how illusions are classified given in Fish (2009: Chapter 6)

If this is classed as any other kind of illusion, then the naïve realist has to give a principled account of at which point on its way to the Moon, does it stop being a veridical perception and start to become an illusion. I predict that any such choice of distance will seem fairly ad hoc without identifying a change in underlying mechanism at a certain point and I am sceptical this is possible because of the gradual progression of the Blipper to the Moon. This defence may also put the naïve realist at the top of a slippery slope, which could end up opening them up to the suggestion that all perceptions are in some way illusory. At any rate this line of defence perhaps misses the point in certain crucial ways, because even if it is classed as an illusion, it still leaves at least two points that need clarifying. These are: how the phenomenal character is constitutively dependent on the object in the past in the veridical case; and also how this now looses this dependence at a certain point away from Earth in the illusory case and exactly what one is aware of now. So given the above, in what follows I will assume that I can focus on analysing the defence that in both cases 'we *veridically* see the objects as they were in the past'.

5.3.2 - The Blipper Screen and Intra-object-time-spread

To illustrate the idea that we don't even see things as they were in the past, I will introduce the Blipper Screen Example. To do this we need to move further away from Earth and introduce more Blippers into the thought experiment:

> Screen of Blippers Example
>
> Imagine a massive screen fifteen times the facing surface area of the Sun, which is covered in Blippers[100]. If we place the screen at roughly the same distance as the Sun, then the differential travel time from the centre of the screen compared with the edges is about one second. Just to start off with we turn the screen on all at the same time, 8 minutes later you will see the Blipper Screen lights turn on in a wave, spreading from the middle of the screen to the edges, over the period of a second. This is because the light

[100] I have calculated these figures using basic trigonometry and kept the features close to what is more easily imaginable; like keeping it close to where the Sun is experienced. Again, whether this set up is actually possible is beside the point, the question is how your perceptual theory deals with it if it were.

> travelling from the edge of the screen takes about 1 second longer to travel to you than that at the middle (c.f. the cinema screen example given above). At the point that you see the screen turning on in a outward wave the Blippers will be switching off in an outward wave, and will be off for the next eight minutes that you experience the screen being fully on.

The point of introducing this example is to illustrate how what we see is actually a conglomerate of any object as it was spread backwards through time. As the Blipper Screen illustrates, even though the Blippers on the screen are turned on simultaneously, this is not what you experience. Your experience is determined by the rate that the light from this turning-on-event reaches you. Explicitly the turning-on-event is simultaneous at the object, but for you it appears to be spread across the period of one second (which is also delayed by 8 minutes). So it seems that it is more correct to say that the way you experience the object is more determined by the simultaneous light that is hitting your eyes, rather than particularly what is happening at the object; although of course this has a distal causal role to play for it to count as a perception in the first place.

Perhaps we can re-phrase this point by suggesting that it is the detected light array that controls what you see and not any actual way that the object is configured at any single time. I will refer to the light detected at the eyes in any one instant as the ‘detected light array’ and abbreviate that to DLAy from here on. Hence what we perceive seems crucially controlled by what the DLAy is telling us rather than necessarily how the object is at any one instant. It seems to me that it is exactly the same process that occurs in earth-bound cases, and hence you experience the world according to the light that is simultaneously reaching your eyes (i.e. The DLAy), rather than what is out there at any one single time. Of course at closer distances this difference will be indiscernible, but the point is that the process is the same in both cases.

This is perhaps where we can appeal to what we have learnt from the previous analysis of the visual buffer. If the DLAy determines the retinotopic way that the eye is stimulated, and this configuration is also passed on and processed in a matching topographic way in the visual buffer, then this must be the information that your eventual visual state is *at least based on*. And as we have seen in the Blipper Screen Example, at large travel times and distances, the state you perceive and the one that is actually out there can be discernibly different. Hence the distribution of the DLAy that is represented in the brain can be discernibly different from that which was emitted simultaneously at the objects. This is one example of how the analysis of the visual buffer can help us here, by showing how our internal representations of external reality may always be slightly mis-matched from what *was* actually out there.

To deny this the naïve realist would have to suggest that the brain somehow corrects for the travel time that the light has undertaken to reach you. As far as I know this claim has not been made and I would suggest that it would be a fairly weak reply without some supporting empirical evidence to motivate it. This could in theory be tested, but I would predict that most people would agree that it is most likely that it would confirm the Blipper Screen Example as I have described it. Therefore I would assume that no one would really want to claim something like the following: that the brain delays the processing of the information reaching it from the centre of the Blipper Screen; in order to match it up with the information reaching it a second later from the edge of the Blipper Screen; because it has somehow worked out they come from the same contemporaneous event and should be represented in experienced as such. So I will assume that the naïve realist must accept that we perceive objects as they were *spread backwards through time* due to both inter-object-time spread and intra-object-time-spread. At least I will assume this until I hear an argument that the brain somehow corrects for this time spread somewhere upstream from the visual buffer. For now I take it that that this defence is unlikely to be forthcoming, although I would be interested to hear it if it is attempted.

This suggests to me that it is therefore more accurate for the naïve realist to change their defence against the Time-spread Argument to one that says something more like the following: we are not aware of objects as they are now, as our naive intuitions tell us; in fact we are not even aware of objects as they were at any one particular time in the past; but rather we are aware of objects as they were spread backwards through time. Hence the defence by Le Morvan is shown to perhaps be *at least* inaccurate and we are now in a position that is even further from what out naïve intuitions tell us. If this interpretation is plausible then this seems to work against the intuitions that motivate Naïve Realism in the first place and hence, seems to me, to weaken their claims in this respect even further. This is because although on the surface it seems like we are just aware of objects immediately and they form part of our experience, by simply applying some basic science and trigonometry, the actual case seems much more complicated and unintuitive. Hence I hope to have explained why I don't find modern Naïve Realism that plausible on at least this count. This potential explanation will be made even more complicated once I have looked at some of the basic metaphysics behind even claiming this modified time-spreading defence.

Before doing that it may be worth developing one other good way of illustrating these points. And that is to imagine that the speed of light was significantly slower so that light took about one second to reach you even from objects held in your hand. You wouldn't really notice the difference for fairly static uniform objects, but for anything moving there would be a very distinct lag between what you saw and where you could feel it was. Rather like the lag between seeing a jet engine and hearing its roar, or the lag between thunder and lightening, but here it is your sense of touch that is working faster than your sense of sight. At this slower speed of light, the time-spread on the Screen of Blippers would be discernable at the scale of a cinema screen too. Hence we can compensate for the relatively fast speed of time at short distances, by challenging the naïve realist to give an explanation of their theory under a thought experiment where there is a much slower speed of light. It seems that all perceptions, even if they were still classed as 'direct', would potentially be non-veridical in this case. If they can't give a principled defence against this possibility then the plausibility of Naïve Realism seems contingently dependent on a rapid speed of light and hence this weakens their position to my mind.

It therefore seems to me, that the fact that we don't notice the time-spread phenomena in earth-bound examples, is just a contingent feature of the rate of our visual discrimination being so much lower (0.1s) then the time-spread discriminations allowed for by the *actual* speed of light (less than 0.01s). We can therefore suggest the reason the Openness and Transparency Intuitions are so compelling is simply a contingent feature of a very fast speed of light, rather than any deeper direct constitutive contact with external reality. At the very least this direct contact is now far more complicated than our intuitions would suggest and is now a fairly involved metaphysical claim. Which brings me to the second part of this analysis: how plausible is to say *we see things as they were spread backwards through time* in the first place? To analyse this we need to look at some basic implications for this approach under different philosophical theories of time.

5.3.3 – Time-spread and theories of time

Very briefly this section will look at the initial plausibility of saying that we are aware of objects as they were spread through the past. To do this I propose that we look at the two main theories of time, namely Presentism and Eternalism, and see how they might help assess this claim.

i) Presentism

First of all I take Presentism to be the most intuitive (common man) approach to time and hence it is prima facia the one that seems the most compatible with the intuitive claims of Naïve Realism; not that much depends on this, as discussed next[101]. According to Presentism (e.g. Markosian 2002) only the present exists ontologically and therefore there are no material objects, or anything of any kind (e.g. matter and energy), that is not constrained within the present time frame. This means that it is metaphysically impossible for objects from the past to be literally a constitutive part of your perception, because there is no physically realised past to play this relational role. The naïve realist may still claim that we can see *now* how things were in the past, and hence appeal to a *perceptual event* that seems to occur entirely in the present, but this does not explain how the non-existant external object in the past is supposed to *literally* constitute the current contours of your experience. So it is hard to understand exactly what the naïve realist thinks we are aware of by appealing to non-existent entities and this claim therefore seems initially implausible given a presentist theory of time.

To put it another way, under Presentism it seems like there is no such thing as objects as they were in the past for your mind to reach out to. And further, even the thought of your mind leaving the present to reach out to the past is probably impossible under this theory of time. This is even if you embrace dualism about the mind and therefore somehow your mind can escape the material shackles of the present, there just doesn't seem to be any material objects to latch on to anywhere else but those in the present. Hence the material

[101] Again this is not to say there aren't good arguments for doubting Presentism, only that I take it that this is what most people on the street would naively agree with if pressed to think about it.

objects that constitute your experience can only be as they are in the present and this removes any relata that that your mind could constitutively perceptually relate you to in the past, which contradicts the defence stated above and hence means that initially Presentism seems incompatible with Naive Realism[102].

It is worth noting that Presentism doesn't claim that objects that were in the past have gone out of existence or have been destroyed and current objects are being perpetually created. Rather that the objects of the past have transformed into objects of the present, which will change into other forms in the future. And these will have continuingly varying properties at these different times according to the laws of physics and entropy. The point is that the objects as they were in the past no longer exist in the form that you experience them as being in *now* and in fact in the case of the Blipper, you never see it as it actually ever is, which is the only state available to reach out to. Hence it seems that under Presentism the objects of our perception can have moved on to instantiate different properties and to exist in different configurations, and there is no past configuration to go back to, let alone one that is spread backwards through time.

ii) Eternalism

The other main option as I see it here, is to be an Eternalist about time (e.g. Sider 2001). The main way of characterising this position is to say that it involves all times actually existing eternally in a *static* 4D block universe. That is the past, present and future all literally ontologically exist, but they are spread out in a way that means they can be represented as in Figure 5.4 below. Here we collapse the three physical dimensions into a 2D plane and plot these against time on the other axis as a 'fourth' dimension:

---

[102] And note I do not deny that you can be in other kinds of relations with objects, such as: remembering them; or thinking about them; or even being taller than them. I am just focusing on my interpretation of the perceptual relation proposed by the naïve realist (c.f. Djukic & Popescu 2003 for another criticism of appealing to other non-perceptual relations to defend a relational *perceptual* theory).

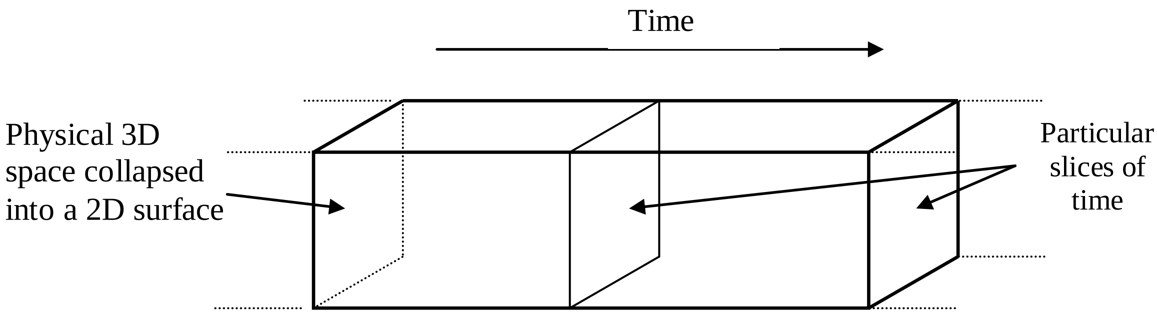


Figure 5.4 – Diagrammatic representation of Eternalism as a static 4D block universe

Although Eternalism is motivated, as I see it, because of concerns to do with physics (e.g. relativity), they have trouble explaining our conscious experience of the forward flow of time and usually have to explain it as merely an illusion. This is because our experience of the flowing passage of time is hard to incorporate into a static universe where nothing even moves or interacts in the way that we normally think it does[103]. Hence if the naïve realist appeals to Eternalism, they immediately trade a strong intuition about perception, for a highly unintuitive one about consciousness and the passage of time, and hence I'm not sure they have made progress in defending their intuitive position. In other words, they might establish that we can have direct unmediated contact with objects as they were spread backwards through time, but concede that the way we are conscious of objects over time is a complete illusion. The net effect perhaps being that we have no better over all epistemological grip on reality than we had before. In addition the indirect realist who is also a presentist could now possibly claim equal rights to intuitive claims, hence nullifying one of the main reasons for favouring Naïve Realism in the first place, as discussed above. So there may at least be an intuitive trade off if this route is taken.

[103] I think it also has many other problems to do with causation and explaining entropy and physics, but these need not concern us here; suffice to say that I prefer Presentism and think I can defend it against the normal raft of objections levelled against it. But that is a very involved process that I can't even summarise here.

Apart from that it is metaphysically complicated, and maybe even more unintuitive, to explain how our perception reaches back, spreading through time, as the illusion of consciousness goes forward, whilst everything is actually frozen in a static universe! So this is not a simple metaphysical claim by any means and one that will need significant unpacking before its plausibility can be assessed. So the idea of our minds reaching backwards through time, in an otherwise static universe, to different time slices of the same object is very involved one to make by Le Morvan and other direct realists. And this makes me very pessimistic about the utility of appealing to this kind of theory of time to defend Naive Realism against the Time-spread Argument. Of course I would be very curious to hear an Eternalism based solution to the Time-spread Argument that still preserves the intuitions of Naive Realism. But until I hear that argument and can assess its plausibility, I feel I am justified in being sceptical about the utility of taking this route for the naïve realist.

This therefore seems to leave no obvious options for the naïve realist to take at this point and hence I see it as an implausible position to hold and this is the main reason why I prefer Generative Approaches. And note I take it as fairly obvious why the Time-spread Argument is not a problem for indirect realist Generative Approaches, since the locally supervenient generation of experience can occur entirely in the present, even though the physical processes that ultimately distally cause it can be spread through the past; I will explain what I mean by this in more detail in Chapter 7.

A quick word might be appropriate here to eliminate appeals to intermediary theories of time such as Pastism: where only the past exists and the present is being continuously created. First of all any theory that posits a static block universe in the past suffers the same fate as Eternalism in terms of awareness reaching back to the past. This is even though it may have less trouble explaining our perception of time flowing in the present. Secondly, the main benefit of Eternalism is that it avoids the ontological excesses of a 'continuously created present', which would bring a staggering amount of matter (a whole universe worth) into existence at every moment. Eternalism does posit a lot of matter into existence (the whole universe for all times), but at least it has always been there and arguably we are just realising how much of it there is. Hence it is not necessarily ontologically profligate, but just committed to a much, much larger ontology in the first place. Therefore I will not deal with these intermediary positions in detail and treat them as less plausible versions of Eternalism. At least I think what I have said above seems applicable to all the most plausible theories of time that I know of.

This therefore concludes the section on Time-spread Arguments against Naïve Realism. The main point I have tried to emphasise is that the normal direct realist claim the: 'we see things now as they were in the past', is inaccurate because it is too simplistic and already seems fairly implausible because of time-spread concerns. This seems particularly acute for naïve realist positions that depend on the object-as-it-was to literally constitute part of your experience. Just the briefest of looks into this area has shown that the Le Morvan style defence is quite a significant metaphysical claim that needs substantial elaboration. It is only at that point that can we tell if this approach still remains plausible and I have given several reasons why I feel justified at present to be sceptical that such an elaboration is possible. To further justify this I have had verbal communication from two current naïve realists that time-spread is still a problem for this kind of approach. That is not to say that I am claiming there is no possible reply, only that I have been given good reason for standing by this argument for now, and why I do not favour theories like Naïve Realism at present. Hence until I get some form of a response to these points, I will carry on as if it is a less plausible approach than a Representational theory of perception. And note I am not claiming to have disproved Naïve Realism, but merely to have shown why I do not favour it and why I feel justified in setting it aside at least for the remainder of the thesis.

There remains a further question as to whether this critique could be applied to other positions that claim to be Selective Approaches and hence direct realist theories. At first glance I would suggest that they would obviously apply to any relational perceptual theory, or one that claims wide supervenience of phenomenal character, where the object is somehow meant to be directly involved in the perceptual state in something like a constitutive role. This is because I would suggest that the arguments stated above question whether it is ever possible to be in any kind of perceptual relation with objects as they were spread backwards through the past. Hence I feel confident that this approach could be applied to at least most relational perceptual theories that claim direct contact with external reality and I take these to be the best candidates for Selective Approaches to perception. Conversely, agreeing with the naïve realists (see quotes above), I would be sceptical that any theory that does not claim such a close relational contact with external objects is properly called a direct realist theory, and we will look at one such theory briefly in Chapter 6. There I will look at the implications of this analysis for Tye's (1995) version of 'Representationalism' and at the same time we will look at some ideas of his that may help analyse the Interpretation Capacity Mechanism. So while I cannot explicitly apply this Time-spread Argument to all direct realist theories here, I remain confident that such an analysis is at least possible (and see below).

On a final note then, I can summarise what I have been aiming for in this chapter. That is to show that Naïve Realism, while admittedly being an intuitive position that satisfies certain epistemological desires, still has significant problems with the Time-spread Argument. To me it seems like the external world just can't play the constitutive role in perception that a Selective Approach might want it to. Hopefully therefore the analysis above has shown why I have focused on this critique rather than others, since I see it as the most compelling reason to abandon Selective Approaches generally. That said, even if a version of a Selective Approach could reply adequately to the Arguments from Illusion and Hallucination *and* the Time-spread Argument, I think it is still explanatorially minimalistic. In other words there is still a significant explanatory debt owed to explain how the *mind reaches out* to make us conscious of external objects and I call this the **Awareness Mystery**. It seems that most Relational perceptual theories just take how we become *aware* of external objects as obvious without feeling any need to explain this because it is so intuitive. However, this mystery is so acute that it leads some researchers to abandon any materialist attempts to explain it:

> 'But the mere fact that [the Theory of Appearing as a relational perceptual theory] takes the appearing relation to be fundamental and irreducible means that it is incompatible with materialism and so will be opposed by materialists for that reason. … I am at a loss to think of what material relation the appearing relation might be, once we reject any reduction of it to the causal chain that eventuates in the perceptual experience. And so [this theory] is opposed to materialism in any form. Since materialism holds no attraction for me, I can cheerfully accept that.' Alston (1999: 196)

So Alston readily accepts that his relational and direct realist perceptual theory cannot be squared with a materialist outlook. However it seems to me that, once we are freed of the confines of materialism, then the mind can spread wherever it likes whether these places still exist physically or not. But this still leaves a really unsatisfying explanatory deficit, and one that is now perhaps out of the realms of empirical study, and therefore it seems that the Awareness Mystery is now even more acute than it was before. For example: how can a non-physical mental state interact with objects as they were in the past, even if they do exist, and why is this relation confined by the speed of light; and how are the contents of the visual buffer accessed to yield this direct immaterial awareness of objects spread backwards through time? To me this would come close to invoking some sort of immaterial mind-beam that is emitted from the brain and can go back instantly in time to put one in touch with the objects as they were. But this is a mysterious entity indeed and hence some elaboration from these kinds of theories about what they have in mind is necessary before we can assess the plausibility of their response to the Awareness Mystery.

So for me the appeal to dualism opens up many more problems than it solves and hence, along with the concerns stated above, it makes me sceptical that any relational direct realist theory is really compatible with a satisfying materialist outlook. I therefore take theories that invoke non-materialist solutions as appeals of last resort and prefer to continue by remaining within a materialist framework and generally abandoning Selective Approaches of perception instead. So until the concerns I have raised here are at least neutralised, then I feel justified in I setting aside all Selective Approaches for now, or certainly at least for the remainder of this thesis. Therefore in the next two chapters we will start to look at Generative Approaches and Representational theories in more detail and specifically look at Tye's version of Representationalism and the Interpretation Capacity Mechanism in Chapter 6 next.

6 - STIm and Tye's Theories of Perception and Imagery

Representationalism and Image Interpretation

'Orgasm is phenomenally **impressive** and there is nothing very impressive about the representational content *that there is an orgasm.* I just expressed it and you just understood it, and nothing phenomenally impressive happened (at least not at my end).' Block (1996: 33)

'The approach I have taken draws heavily on the views of both [depictivism] and [propositionalism], but is not properly classifiable as falling into either camp. The problem with the theories I have rejected, as I see it, is not that they are completely wrong-headed but rather that they are only partly right. In some respects mental images are a little like pictures, but in others they are like descriptions. The truth about images, I suggest, is that they are a mixed breed.' Tye (1991: 102 – names of theories changed to match the equivalent terms used here)

In the last chapter I introduced and rejected Naïve Realism as a paradigmatic example of a Selective Approach. I did this by relying mainly on the Time-spread Argument, but I also explained why I find other motivations for defending Naive Realism uncompelling. I suggested that these arguments might generalise to other forms of Selective and Relational Approaches to perception, although I did not have time to go into this latter claim in much detail. Getting to this position now allows me to focus on more Representational, and possibly Generative Approaches, in the remainder of the thesis and the full positive theory that I favour will be developed along these lines in the next chapter. Before I can do that there are two more features that I think I need to address, and both of these involve analysing two different aspects of Tye's (1991, 1995) work on the relation between perception and imagery. Hence both of these analyses will contribute to refining and defending my final theory of STIm as an offline simulation of visual perception. Hence in this chapter I will start by analysing whether Tye's (1995, 2000) Representational PANIC theory of consciousness is best interpreted as a Selective or Generative Approach, especially in terms of visual perception. I will then go on to analyse some earlier ideas by Tye (1991), to do with how visual images come to be *interpreted*. This will act as a way to develop some ideas of my own about the Interpretation Capacity on the output side of Figure 4.3 as discussed in Chapter 4. I will now say a little more about each part of this chapter in terms of an introduction before starting the analysis proper.

As mentioned briefly in Chapter 1, I defend an approach that is very Tye-like at least at the level of symbol manipulation in related areas of the brain. This is because Tye (1995) also suggests the visual buffer is utilised in both vision and visual imagery and they are both based on a similar kinds of hybrid depictive representational processes. For Tye this means they are both based on 'interpreted symbol filled arrays' in the visual buffer and I repeat the most pertinent quote from Tye to illustrate this point; and note that this is the same quote that was first given in Section 1.3:

> 'There is strong evidence that images and visual percepts share a medium that has been called the "visual buffer". This medium is functional: it consists of a large number of cells, each of which is dedicated to representing, when filled, a tiny patch of surface at a particular location in the visual field. For visual percepts and afterimages, the visual buffer is normally filled by processes that operate on information contained in the light striking the eyes. For mental images (other than afterimages), the visual buffer is filled by generational processes that act on information stored in long-term memory about the appearance of objects and their spatial structure. Images and percepts, I have argued elsewhere, are *interpreted, symbol filled patterns of cells in the visual buffer*.' Tye (1995: 122: Box 4.7 – footnotes removed, emphasis mine)

I will explain what this may mean in more detail below in the **first** part of this chapter. There I will suggest that adopting Tye's underlying structure still seems to be currently viable and is consistent with much that I have already discussed in terms of defending the Kosslyn Model in Chapters 3 & 4. This is what makes Tye's PANIC theory, and what he thinks happens next, very relevant to this project. However it is at the point of discussing how these underlying symbol filled arrays lead to full phenomenal consciousness that I will mainly disagree with Tye. This is because I will suggest that his form of reductive[104] Representationalism is best seen as a Selective Approach and hence suffers many of the same problems I have identified in the last chapter. I will argue that, in my opinion, he fails to explain adequately how accessing these symbol filled arrays in the visual buffer leads to phenomenal consciousness. So while accepting much of his underlying PANIC structure, I will slightly modify the top level of his representational theory in favour of a less representationally reductive approach (and these terms will be explained further in what follows).

In the **second** half of this chapter I will focus on what Tye might mean by the idea that these symbol filled arrays need to be 'interpreted' to become mental images. Although, as indicated above, he includes this as a qualification in his later (1995) full PANIC theory, he says little about it there, and we therefore need to look at an earlier book by him called 'The Imagery Debate' (1991) to explain this qualification[105]. Regarding this issue I will again argue against Tye and suggest that un-interpreted imagery can occur, although sometimes I think it needs to be 'labelled' appropriately so that it can be used for a willed episode of problem solving. However this still allows seemingly spontaneous and abstract un-interpreted and unlabelled images to occur, and I think this reflects our phenomenological experience of imagery much closer. We will therefore need to look at the distinctions I will develop between 'image-labelling' and 'image-examination' and analysing this will form the bulk of my criticism of Tye's (1991) interpretation-based approach.

---

[104] I will explain this 'reductive' distinction in what follows, but generally it refers to the way he reduces phenomenal character to representational content without the need for any extra explanatory step or mechanism. I will defend a non-reductive form of Representationalism in the next chapter that accepts some form of extra generative step is necessary.

[105] Note that I have used other arguments from this text to refine the distinctions about hybrid depictive representations as discussed in Chapter 4. Here we will look at a different aspect of this book.

Discussing these ideas will also allow me to suggest that this is how visual images become *labelled* with a temporal context and this is what allows them to represent events that may occur in the future, or could have occurred in the past. This section will therefore constitutes my main arguments for how the temporal and spatial context becomes associated with STIm. Recall I have already discussed in Chapter 4 how the *spatial layout* of features that are contained within the imagined scene are controlled by integrating it with information processed in the SPP stream. So in addition to the way that objects are spatially laid out in the imagined scene, I will also discuss here how this scene can also be labelled according to different spatial contexts. For instance, you can visually imagine the spatial layout of your room being located in the jungle or the desert, even though the actual distribution of the objects in your room can be the same in each case. This allows me to distinguish between the *spatial layout* of an image, which is determined by how the objects in it are distributed in apparent 3D space, and the *spatial location* of that scene, which determines what geographic location this is set in and under which you consider it. But note that this spatial context can be left unspecified if not relevant and can be an imaginary location too. Hence by discussing this labelling and interpretation process I hope to develop a way to explain our full capacity to imagine different arrangements of objects, under different temporal contexts, and in potentially different spatial locations. This will form the platform for explaining my full theory of STIm in Chapter 7.

So specifically this chapter will focus on two relevant aspects of Tye's work and develop them as a platform for the concluding chapter to follow. To recap, it will **firstly** argue that Tye's (1995) PANIC theory can be interpreted as a kind of Selective Approach and if so inherits some of the problems discussed in the last section. I will also briefly discuss some other problems with his PANIC theory as I see it and link this with more recent work by him (e.g. Tye 2009). And **secondly**, this chapter will develop Tye's (1991) idea that the symbol filled arrays in the visual buffer need to be *interpreted* to become mental images. I will instead suggest an alternative approach, which argues that they only need to be *labelled* appropriately if they are to be used in a willed episode of imagery. This labelling step will also accommodate the idea that they can be labelled according to an alternate temporal context and/or spatial location. This will provide the platform I require for developing a full account of STIm based on a Generative and Representational Approach in the next chapter. So we will start now by analysing Tye's (1995) PANIC theory and why I think it is best seen as a specific variety of a Selective Approach.

6.1 – Tye's PANIC Theory of perception and imagery?

In this section I will first briefly introduce Tye's (1995, 2000) PANIC theory of phenomenal consciousness and then I will criticise it. As mentioned above, I will mainly accept Tye's account of the underlying structure of how phenomenal consciousness is supported (or underpinned). My criticism will mainly come in at the level of how Tye explains how phenomenal consciousness relates to this underlying physical framework. I will suggest that he provides no adequate account for this unless he is suggesting a version of a Selective Approach, in which case he may be subjected to the arguments I made against these kinds of theories in the last section. Hence this first part of the chapter will be split into two main phases: firstly I will simply introduce his PANIC theory, and then I will argue that it might best be seen as a Selective Approach, and if so then I can reject based on the Time-spread Argument as described in Section 5.3.

6.1.1 – Tye's PANIC theory – the underlying structure

In my opinion the best bit of Tye's (1995) PANIC theory is his explanation of the basic underlying structure of phenomenal consciousness, which is heavily empirically guided and remains plausible even today. This relies mainly on the uncontroversial suggestion that sense organs, such as the eyes, ears and skin, collect information from the environment, which gets processed via *early* raw sensory arrays, into *final* sensory arrays (FSAs) in the brain. This involves a relatively straightforward process of detecting incoming signals in transducers in the sense organs and relaying these to neural network arrays in the brain for further processing. Hence these FSAs could be said to represent the information detected from the body and the environment via the sense organs. A detailed example of this causal process in terms of vision has been described in Chapters 3 & 4, while defending the Kosslyn Model. And some of Tye's earlier (1991) comments about how these representations are utilised by the visual buffer have been discussed and incorporated there. This part of Tye's PANIC theory therefore currently remains highly plausible in my opinion, even given the updated version of the Kosslyn Model discussed earlier; indeed this part of his work has been much of the inspiration for the direction this thesis has taken. Hence I will now introduce Tye's next step in just enough detail, so I can analyse his transition from the FSAs to full phenomenal consciousness.

To that end it should be noted that Tye's full PANIC theory is aimed at explaining all forms of phenomenal consciousness, and hence my focus on visual perception in this context, is just a specific subset of his full theory. Hence it seems possible that equivalents of the 'visual buffer' could be invoked for other sensory modalities. For example, it is known that the somatosensory cortex is subdivided into areas that correspond to skin on different parts of the body, with areas that are more sensitive having more volume of cortex devoted to them, as depicted in the diagram below:

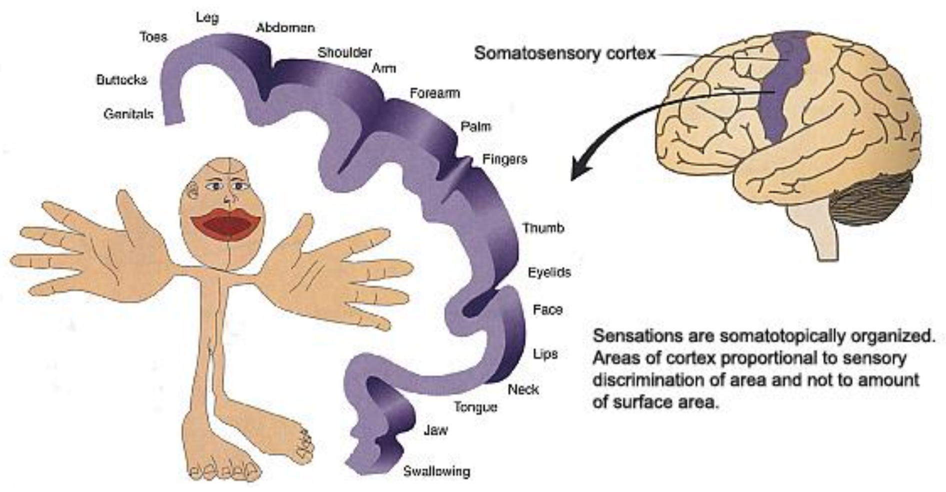


Figure 6.1 –Divisions of the somatosensory cortex that associate pain sensations to areas of the body. On the right is its location in the brain; on the left is an illustration of the relative amount of brain matter devoted to each region of the body. Note for example how the hands and lips have a very large amount of brain matter devoted to them compared with their relative actual size (diagram adapted from Carlson 1994)

So Tye thinks a similar thing can be said for other sense modalities and this purportedly allows him to explain how more general things like orgasms and moods can represent things going on within the environment and ourselves. I do not propose to go into whether these arguments are successful or not (although I think they largely are at the symbolic level), as this is not the area I have in contention with these theories[106]. Hence I will focus simply on visual perception from here on, but be aware that in some of the quotes below, Tye is talking more generally too. Although having said that he does also tend to focus on the visual modality, as do most other perceptual theories in the literature.

In terms of vision Tye (1995) recognises that there are subsidiary processing steps that subserve the upstream sensory modules, like edge detection and colour processing. These in turn yield as its output ‘a unified representation of the entire visual field’ (pg. 138). He equates this upstream sensory module in vision with Kosslyn’s Visual Buffer as discussed

[106] But see Block (1996, 1998, 2003: and see opening quote), Tye (1996, 2003, 2005), Rey (1998), Alston (2005) and Crane (2001: 78) for a sample of other research on these issues.

above. Here is how Tye explains the steps going from the FSAs in the sensory modules in the brain (i.e. the visual buffer) to full blown phenomenal experience:

> ‘Representations occurring within the modules supply information the creature needs to construct or generate sensory representations, but they are not themselves sensory. Phenomenal content, in my view, is not a feature of any of the representations occurring *within* the sensory modules. As I noted in the last chapter, *experience and feeling arise at the level of the outputs from the sensory modules and the inputs to a cognitive system.* It is here that phenomenal content is found.
>
> Sensory representations (viewed in the above way) represent either internal or external physical items. Bodily sensations represent internal bodily changes. They are directly tuned to such changes (in optimal conditions). Likewise emotions and moods. In the case of perceptual experiences, the items sensorily represented are external environmental states or features.
>
> Phenomenal content, I maintain, is content that is appropriately poised for use by the cognitive system, content that is abstract and nonconceptual. I call this the PANIC theory of phenomenal character: phenomenal character is one and the same as Poised Abstract Nonconceptual Intentional Content. …It follows that representations that differ in their PANICs differ in their phenomenal character, and representations that are alike with respect to their PANICs are alike in their phenomenal character.’ Tye (1995: 137 – referred to as **Main Quote** in what follows - my emphasis in quote)

I will not discuss the ‘in optimal conditions’ clause here, because this qualification is not relevant to my arguments in this context, although it becomes quite crucial later for his responses to Inverted Earth arguments (Tye 2000: Chapter 6). For now we can see that for Tye, ‘phenomenal content’ is just a form of ‘intentional content’ because the FSAs, which are the basis for phenomenal content, are all *about* (or they represent) something in our surrounding environment or bodies, and hence they are intentional.

Hence Tye maintains that: ‘Phenomenal character (or what it is like) is one and the same as a certain sort of intentional content’ [pg.137 – Box.5.2]. He thereby equates phenomenal character with Poised Abstract Nonconceptual Intentional Content (or PANIC for short): the **poised** indicates that it is not always accessed by the Cognitive System to form beliefs and desires but only stands ready to be; the **abstract** indicates that the state does not have to have a real concrete object as its focus, it can be real or non-existant (e.g. hallucinations); the **nonconceptual** indicates that the subject need not have matching concepts for the general features entering into these contents; and the **intentional content** indicates that it is about or directed at (or represents) something.

He ends his section on PANIC by re-stating his reduction of phenomenal character to representational content, which presumably is aimed mainly at qualia theorists who might argue that phenomenal character can out-stretch representational content:

> ‘So that is what what it is like is really like. It is PANIC. Philosophers who have attempted to draw a sharp distinction between the representational aspects of our mental lives and their phenomenal or subjective or felt aspects are mistaken. The latter form a subclass of the former, like it or not.’ Tye (1995: 143)

This is why, in line with other prominent interpretations available in the literature (e.g. Lycan 2006 and Chalmers 2004), I think Tye’s approach is referred to as a ‘representationally reductive’ position. This is because it reduces phenomenal character to representational content without remainder. That is, it suggests there is no need to appeal to anything ontologically new at this point, such as qualia, or any immaterial sense-data, for instance.

It is worth noting this kind of reduction can be importantly differentiated from reduction to the physical because even representationally non-reductive theories can claim to be compatible with materialism. They must however then just claim that anything else involved in their theory is also physical. Hence this allows a space for those who favour a less representationally reductive approach, but still want to remain broadly monist about substance, even though materialist non-reductive representationalism may seem to be at first sight a bit less intuitive[107]. This latter approach is the kind of position I favour and these issues will be discussed further in the next chapter.

Also note that this 'reductive vs. non-reductive representationalism' distinction is different to the 'strong vs. weak representationalism' distinctions in the literature. This is because strong representationalism only suggests that representation of a certain kind suffices for phenomenal character. Conversely weak representationalism suggests only that phenomenal character has some representational content, and hence neither say anything specifically about a reduction to representational contents. This point is illustrated by noting that strong representationalism can come in reductive and non-reductive forms, where the latter could appeal to qualia in one form or another. This way of splitting the 'Representational' field is in line with how I think Crane (2006) would crucially distinguish it from Relational theories, as discussed in Chapter 5. And it is also probably why I think Crane takes the differences between different, broadly speaking, representational theories (e.g. Tye and certain qualia theories) as only being a minor crevasse, when compared with the large chasm between Representational vs. Relational theories[108].

---

[107] It is less intuitive because Reductive Representationalism seems to rely in all its stages on elements that are more obviously physical, and this may have been one of the main reasons for motivating it initially. See MacPherson (2003: 56 & footnote 27) and Lycan (2006) for an expression of this sentiment about the work of Dretske (1995) and Tye (1995).

[108] Here he is rejecting Block's (1996: 19) claim that the dispute between Representationalism and Qualia Theory is the 'greatest chasm in the philosophy of mind'.

This categorisation also possibly matches up with that indicated by certain other researchers who may only reluctantly accept being categorised as qualia theorists. For instance, in the introduction to a recent collection on ‘qualia theories’, Wright (2008: 2) laments the exclusive use of the term “Representationalism” for Tye-style reductively representational theories. When actually historically older representational theories were sympathetic to qualia theories. For example, Wright’s (1993) first edited collection was called “New Representationalisms” and according to him those articles all ‘espoused’ qualia. Hence Wright has abandoned the use of the term “Representationalism” in favour of defending a broadly construed “Qualia Theory”. However with the *possible implicit* recent rejection by Tye (2009) of his PANIC theory because it appeals to phenomenal concepts (as briefly mentioned in Section 1.4 footnote 22 and see below), and the complications involved in using the now more technical term ‘qualia’, I would suggest that perhaps abandoning this useful term may be too hasty. Perhaps it is better to suggest that Tye’s PANIC theory and certain other qualia and/or intentional theories are both varieties of Representationalism but are crucially different in how they expand on that. I think this is best reflected in referring to Tye’s approach as Reductive Representationalism, and to theories that are more sympathetic to Generative Approaches, as non-Reductive Representationalism. This is the convention I will adhere to in what follows where I will criticise Tye’s Reductive Representationalism and this will allow me to defend a non-Reductive Representationalist theory in the next chapter.

However, it is worth noting that even given my interpretation of the way this field is divided, which is in turn based on the interpretation of leading researchers within this same field, this is a complex and evolving matter and neat categorisation may not always be possible. Hence I will still try to stick to the classifications that are most important to my analysis of STIm in the context of this thesis, and that is mainly the distinction between

Selective and Generative Approaches. So in my opinion the most important next step is to analyse how a Reductive Representational approach like Tye's PANIC theory, attempt to explain phenomenal character and this is where I think he can be interpreted as possibly condoning a Selective Approach. But before I can justify this conclusion I need to analyse the steps in Tye's theory between the FSAs and the eventual experience of phenomenal character.

6.1.2 – Problems for Tye's PANIC theory

So now that I explained what I mean by Reductive Representationalism we can return to specifically analysing Tye's PANIC theory. Given the Main Quote from Tye (1995: 137) above and especially the part I have emphasised with italics, he seems to suggests that it is somewhere between the output of the final sensory modules or arrays (the FSAs like the visual buffer) and the input to a Cognitive System that phenomenal content is found and experiences and feelings arise. It is worth noting that throughout his body of work he says practically nothing about what he thinks the 'Cognitive System' actually is, apart from containing a mechanism for forming beliefs and desires based on some of our percepts. He also fails to address just what happens at the output stage from the visual buffer to a stage that it can be accessed by the Cognitive System appropriately. Hence there appears to be no explanation of the way these FSAs output their information in order to make the symbolic representations available to phenomenal conscious and hence poised for the Cognitive System to access. And therefore it seems to me like he just leaps across the signal to phenomenal gap without explaining how it's done. This is especially odd after making the problem of mechanism at the gap so clear earlier in his book:

> 'How can electrical activity in dull grey matter generate feeling and experience? What is wanted here is a description of *mechanism* that underlies the generation of perspectival entities from nonperspectival ones and that closes the enormous gap

we intuitively feel between the two. This is the problem of mechanism.' Tye (1995: 43)

I refer to this as the Generation Mystery and it is something any Generative Approach to perception must accept exists. That is there remains a question as to how the brain generates phenomenal character from these simple physical and symbol manipulation processes such as neurons firing in arrays in the brain like the visual buffer. And note this is different to the Awareness Mystery, which I introduced at the end of Chapter 5, which questions how the brain supports us being directly aware of the shapes and colours of external objects. So in the former case the brain generates phenomenal character itself and in the latter it merely supports awareness of external objects and their properties, and Tye doesn't seem to give an explanation of how either is achieved or even that it needs a separate explanation.

This lack of detail in Tye's PANIC theory has led Kriegel (2002a) to suggest that none of the individual PANIC criteria are enough to discriminate phenomenal from non-phenomenal mental states. This is because it seems that non-phenomenal mental representations can be: abstract and non-conceptual and have intentional contents, just by representing a certain kind of external object that may or may not actually exist via an FSA. Kriegel then argues that *poisedness,* the only feature that is now left, is not a difference in phenomenal content but rather a *vehicular property* in the different functional role it can play. So the fact that it can be experienced phenomenally is not about a difference in its phenomenal contents (what it represents) but rather how it functions: i.e. whether it is poised or not. And this might risk collapsing a representationally reductive theory into a functionalist theory, and functionalist theories are mainly criticised for their lack of explanation of the phenomenal[109].

So it seems that Tye's identification of phenomenal character with PANIC leaves what happens between the output from the visual buffer and phenomenal experience seriously under described. Given the functional interpretation above it seems that just having these representations appropriately poised doesn't explain how potentially non-phenomenal representational content becomes phenomenal content and hence phenomenal character. So we must perhaps look elsewhere for an explanation of what is happening in the case of perception and perhaps try and identify an *extra step* that helps explain how Tye thinks phenomenal character arises. Related to this I will analyse and reject the idea that an interpretation process that Tye could appeal to might be a suitable candidate in Section 6.2 next.

Before doing that however, it think it would be appropriate to investigate whether, given the above, Tye's PANIC theory can be interpreted as being a Generative or Selective Approach, and this can then be used to relate his theory to my analysis of STIm. That is, I will analyse whether in perception the extra step might involve: *either* reaching out to put one directly in touch with external objects and their properties, as a Selective Approach might suggest; *or* would it involve the brain generating phenomenal character based on the contents of the visual buffer, as a Generative Approach might suggest[110]. It seems to me that both of these steps could potentially take the outputs of the visual buffer and make them phenomenally available for the Cognitive System to utilise in further thought

[109] C.f. Block (1978, 1980), and Churchland (2005) for a review of 40 years of Functionalism.
[110] I will assume here that this exhausts the possibilities open to Tye at this point given that he needs to explain where the phenomenal character of our experience comes from in some materialist way: i.e. he is not eliminativist about it and hence it does need an explanation of some kind.

processing. That is after this extra step the phenomenal character can be poised and one can attend to it or not and hence form beliefs and desires about it or not.

Given PANIC theory's representationally reductive nature I would suggest that it is not best seen as a Generative Approach, as this step would then need to be explained separately, and Tye explicitly denies in the quotes above that anything else is necessary. He also seems to want to avoid at all costs introducing any *potentially* immaterial intermediary by which we experience external objects indirectly[111]. So I do not favour interpreting his position as a Generative Approach.

Conversely, given Tye's extensive appeals to the Transparency Intuition, I would suggest his theory is most naturally interpreted as a version of a Selective Approach. He seems to expressly suggest that in perceptual states it is the properties of the objects themselves that we are aware of directly. Consider the following two quotes from sections of his work specifically appealing to the Transparency Intuition:

> 'The lesson of the problem of transparency is that *phenomenology ain't in the head*… to discover what it's like, you need to look outside the head to what brain states represent. Phenomenology is, in this way, externally based. So systems that are internally physically identical do not *have* to be phenomenally identical.' Tye (1995: 151)

> 'An explanation is needed of why the phenomenal character of visual experiences is sensitive in this way to surface qualities – qualities that, if they are anything at all, are qualities of surfaces experienced. … the explanation surely is that the phenomenal character i*nvolves* the surface qualities of which the subject of the visual experience is directly aware – that these qualities at least partly *constitute* phenomenal character… The best hypothesis, I suggest, is that visual phenomenal character is representational content of a certain sort- content into which certain external qualities enter.' Tye (2000: 48)

[111] C.f. his pervasive, and some would say far too brief, dismissals of Sense-Data theories: e.g. Tye (2000: 45-46) and Maund (2005: 40) for a critique.

The combination of these quotes together with his claims to being a direct realist theory, suggest to me that Tye takes the Transparency Intuition to favour the idea that it is the qualities of the surfaces of the objects themselves that at least partly *constitute* phenomenal character and which we experience directly by representing them in the way described above. For example, in the first quote he seems to be suggesting the phenomenal character of our perceptual experiences has a wide supervenience base and hence phenomenology involves external objects as a Selective Approach might suggest. To deny that internally physically identical duplicates must be phenomenally identical also seems to favour this interpretation and goes against the BIV intuition that a Generative Approach would accept[112] (as discussed previously). In the second quote he seems to suggest that it is the intrinsic qualities of the objects themselves that we are aware of and in Tye (2000) he argues that colour is an intrinsic physical property of the external objects[113]. And note his suggestion, which echoes that of the Naive Realists' (c.f. Section 5.1), that the qualities of the objects themselves at least partly *constitute* the character of one's conscious experience. This therefore seems like a theory that suggests it is just the objects themselves, and their respective surface qualities, that we are aware of in visual perception, and hence one I would interpret as best falling under the category of a Selective Approach to perception as defined here.

---

[112] This quote seems to deny internalism about phenomenal character and Tye seems to still favour externalism in his most recent work (2009: Chapter 8). For an internalist thesis about BIVs see Horgan et al (2004).

[113] It is a currently debated point as to whether colour is actually an intrinsic property of an object at all as Tye's theory necessitates (as does Naïve Realism). Hence Tye (2000: Chapter 7) spends some time defending the position of Colour Realism, but as Lycan (2001: 20) notes: 'Of course, color realism has been a minority position in the history of philosophy, so this must be counted as a liability of [a reductive] Representational theory.' There remain serious problems with Colour Realism such as the problem of metamers and colour constancy: see the following for more on this debate: Hardin (2008), Byrne and Hilbert (2003), McLaughlin (2003) and Wright (2003). I also have an original criticism against Colour Realism based on the contents of the visual buffer that I hope to develop in future. Basically it states that we are not aware of the light that is reflected at the object, but rather what arrives at our eyes and the two are rarely, if ever, the same due to differential wavelength absorption on propagation. Hence we cannot be aware of the colours that external objects are, even if they are coloured. I see this as a version of the Awareness Mystery because it is still a mystery to me how we are aware of the colours objects have even if they do have them.

This is because it seems that all the brain now does is select which external properties one is directly aware of by representing them appropriately, rather than locally generating these phenomenal states. So it seems this selective process would now need to be explained as occurring via an underlying representational method as opposed to a more obviously relational scheme such as those invoked by Naïve Realism for example. Hence I suggest that under Tye's PANIC theory our representational content puts us in touch directly with external reality, and hence PANIC theory is best classified as a Selective Approach in this respect.

If the above interpretation is correct then Tye's PANIC theory can be criticised using the Time-spread Argument as described in the last chapter. This is because the representational contents formed based on the processed data in the visual buffer, will be delimited by the current information detected at the eyes (the DLAy), which is also in accordance with Tye's theory. So at large distances, in an environment with rapidly changing properties (say a Blipper Screen), the representational contents may not now correspond with what is actually out there in the external world at present. If the environment is not instantiating the right surface qualities then these cannot *partially constitute* what we experience, and this leaves no explanation of where the phenomenal character of our experience comes from under a Selective Approach. To put it another way, what the PANIC states represent as being out there will be significantly out of phase with what is actually in the external world in these cases. And since it is the intrinsic physical surface qualities of the objects (i.e. colours) that we are aware of, as properties of the objects, if these objects no longer exist, then neither to these properties[114]. Hence if Tye's theory is really suggesting a

[114] And note this prevents Tye appealing to universal properties that do not go out of existence with the object, since these colour properties are supposed to be intrinsic properties of the objects: namely surface

Selective Approach to perception then it seems to me to also fall foul of the Time-spread Argument. This is because I don't think any of the PANIC features can explain how we become aware in the present, of objects as they were spread out backwards through the past. That is in any better way than any of the more obviously relational perceptual theories discussed in Chapter 5; or at least this is what I will assume here until I hear a reductive representationalists reply to this criticism.

The problem for Tye as I see it, is that he doesn't provide a satisfactory account of either the Awareness Mystery or the Generation Mystery. Instead he suggests that just by representing something based on the outputs of the FSAs in the brain, which are in turn poised in the right way to be accessed by the 'Cognitive System', that somehow we become at directly aware of external objects and their surface qualities. And further this is even true when those objects and properties are spread backwards through the past and hence may have drastically changed the surface qualities that they are instantiating during the time it takes the reflected light to reach our eyes. A mystery still remains as to how this one-way causal system (object-to-eyes-to-brain) somehow makes us directly aware of these physically and temporally distant objects as they were spread backwards through the past, some of which may no longer exist. So just saying that the phenomenal character just is the same as the representational content does not explain these features, because the representational content as far as I can tell, is just based on some neural networks in the brain (the visual buffer) that is output and appropriately poised to be accessed by some other part of the brain (the so called 'Cognitive System'). Hence it seems like there still remains an Awareness Mystery of how this very plausible underlying system makes us

spectral reflectances (see Tye 2000: Chapter 7). And see Dunn's (2008: Section 5) analysis of the implausibility of appealing to uninstantiated universals (especially for a materialist).

phenomenally conscious in a way that seems direct and partly constituted by the object and its surface qualities[115].

As discussed briefly in Section 1.4 (c.f. footnote 22), normally Tye (2000: Chapter 2) would appeal to 'phenomenal concepts' to explain why any remaining apparent explanatory gap is merely due to a cognitive illusion. However Tye's recent book called 'Consciousness Revisited', has rejected the plausibility of this appeal to phenomenal concepts and therefore it seems that this move is not open to him anymore (see Tye 2009: and especially Chapter 3). My interpretation of his new position is that he has now moved to a more obviously Selective Approach, based on knowledge by acquaintance, and hence the arguments above and in the last chapter will be even more pertinent. He certainly seems to still be committed to the Transparency Intuition (pg. Xiii), and repeating my agreement with Kennedy (2009: abstract) on this issue from Section 5.1, I would suggest this intuition is most compatible with Naïve Realism and Relational perceptual theories, as opposed to Reductively Representational theories[116].

So, given the above, I feel I can reject Tye's PANIC theory of conscious perception for being overly representationally reductive and as long as my interpretation of it as a Selective Approach is appropriate. If however Tye had moved to a more Generative Approach then I think this would resemble something more like the theory I will defend in the next chapter. So in many ways my theory represents an extension of Tye's underlying structure, but where the Transparency Intuition is accepted as misleading due to the Time-

[115] And note this problem occurs irrespective of the underlying representational system you adhere to: i.e. propositionalism or depictivism. For example, how does a propositional representation in the brain lead to phenomenal consciousness, especially if the external world is not instantiating that property anymore.

[116] It should be noted however that the interpretation above of his very recent work is relatively tentative and I eagerly await the first reviews and replies to his new work to test this hypothesis further. It suffices for me to say at this point that his new book doesn't seem to address any of the problems I identify here as far as I can tell and if anything I think it perhaps makes them more acute.

spread Argument and hence as a result the brain is cited as needing to generate phenomenal character itself. This would also entail that phenomenal character cannot be reduced to representational contents alone as Tye suggests, and some further generative step is also required. This involves suggesting that the brain can generate phenomenal character in all cases, which is in line with what I think the Time-spread Argument compels us to accept because the external world does not have the right features to do this. It also involves accepting that there is still a Generation Mystery and this will be discussed further in what follows. So this is the non-Reductively Representationalist direction I have adopted, based on the underlying structure of Tye's PANIC theory, and this will be developed and defended further next, in Chapter 7. Before doing that I will look at the other relevant aspect of Tye's work I mentioned above, and that is whether symbol filled arrays necessarily need to be 'interpreted' to become visual images.

6.2 – The Interpreting and Labelling of Visual Images

In Chapter 5 of Tye (1991), he develops what I will call an 'interpretation-based' approach to mental imagery as summarised in the quote below:

> 'The arrays, in and of themselves, merely represent surfaces and locations as seen from a particular viewpoint. Mental images result only via specific interpretations of the arrays. For example, an array representing a horse-like appearance becomes an image of a horse via the imager's interpreting it as representing a horse. I shall assume that it is correct to view thought generally as involving linguistic representations in some neuronal code, that is representations have combinational syntax and semantics. Hence, I take the interpretation here to require the appending of an inner sentence with the content "This represents an F" to the array.' Tye (1991: 94)

I take this quote to suggest that the information stored as hybrid depictive representations in the visual buffer array need to be interpreted for them to become experienced images. There are a number of reasons I think this approach should be resisted and the extent of the interpretation phase limited to only certain specific episodes of imagery. Therefore in the first part of what follows, I will I suggest that the Tye's 'interpretation' step is only required in a consciously willed episode of imagery. It just so happens that these occasions are the ones that I am most interested in analysing in this thesis. I will then suggest that this process might better referred to as a 'labelling' process and the term interpretation reserved for the eventual *examination* of the visual image. This 'labelling' step also provides a good way of getting to my final conclusions about how STIm gets allocated its spatial and temporal context and hence forms a vital step in my arguments. Therefore I will start by arguing against Tye's necessary interpretation step in the next sub-section, before developing my own theory with its own 'labelling' terminology.

6.2.1 - Arguments against the necessary interpretation step

The **first** problem I have with Tye's interpretation-based approach to imagery, is that it seems the possible range of our phenomenal mental imagery experiences far out-runs the capacity of any linguistic representation to *interpret* it[117]. For example abstract rapidly changing shapes, amorphous blobs, and other seemingly un-describable imagistic experiences, would all seem to be dissociated from any real ability of an interpretation sentence to capture the detail of their phenomenology; that is, certainly in enough detail to mark the difference between one amorphous blob and another. It therefore just doesn't seem to be the right kind of thing to make the difference between a symbol filled array and the equivalent experienced mental image.

This is particularly pertinent when we recall that Tye invokes the same procedure for perception, as this quote suggests: 'Images and percepts, I have argued elsewhere [i.e. Tye 1991], are interpreted, symbol filled patterns in the visual buffer' [Tye 1995: 122: Box 4.7

[117] A similar point is made by Martin (1992) in terms of experience out-stripping our possession of intentional concepts.

– also quoted above]. To illustrate this point in terms of perception, take for example a perceptual experience of a complex fractal pattern on a poster. What difference exactly is the ability to append the perceptual experience with the interpretation - “This represents a fractal pattern” – meant to make. Certainly there is a difference in understanding when looking at this poster if one possesses the ‘fractal pattern’ concept. One can’t really *see it as a fractal pattern* and refer to it as such if one didn’t have that concept to deploy. But possessing this concept or being able to mentally interpret it as such, just doesn’t really seem essential in order to change it from an array to a complex phenomenal experience. Conversely, is it a consequence of this approach, that if you lack such a capacity to form this inner sentence, then you cannot see the picture? I would suggest that we could all see fractal patterns before we ever learnt the concept or had a suitable linguistic mental representation under which to bring them.

Additionally this interpretation step requirement also seems to be at odds with his general PANIC theory where these arrays can be non-conceptual and abstract, and only need to be poised in the appropriate manner so they can be accessed by the ‘Cognitive System’ in order to produce further beliefs and desires. Unless he links this interpretation requirement with some kind of limit to the poisedness criteria, it seems that just the information output from these arrays should be enough to support phenomenal experience without them also needing to be interpreted. After all, they do contain all the symbolic information one would need to create an image (surface colours and their locations) irrespective of whether it is also interpreted or not[118]. It seems that this interpretation step is more to do with the Cognitive System, which includes a belief and desire forming mechanism, bringing these experiences under concepts, like the belief that “This horse-shaped object in front of me

---

[118] See also my criticism of propositional only theories in Section 3.4 where I argued that they might also have to appeal to padded arrays of propositions in order to describe every pixel of the visual field in rich enough detail.

represents a horse"; or more simply "This represents a horse". And recall that according to Tye's PANIC theory this is not done for all phenomenal experiences we have but only stands poised to be available for beliefs and desires to be generated from them by the 'Cognitive System'.

I would therefore suggest that we could have imagistic sensory experiences of horse-like shapes without necessarily appending an interpretation sentence to it of the form: 'This represents a horse'. This would involve just having horse-shaped experiences without bringing them under any particular concepts or linguistic mental representation. This seems like the right way to go about it because all the information required to form a mental image is presumably stored in the visual buffer, and this just needs to get converted so that some awareness of what is represented occurs. This could just be some automatic conversion mechanism (maybe analogous to painting-by-numbers), and it seems unclear why appending this with an interpretation changes it from mere symbolic representations of 'surfaces and locations as seen from a particular viewpoint' to an image of a horse. Surely we can have a mental image of something horse shaped without ever previously experiencing or learning about horses, or even having some internal linguistic mental representation of a horse. On the other hand, we can grant that one can't treat this image as of a 'horse' unless one has the right concepts to bring the image under.

The **second** main problem I have with this approach is that there seems to be many examples of un-interpreted phenomenal experiences that we may barely even attend to or particularly notice. In terms of mental imagery, examples might include an unremarkable daydreaming episode or simply a seemingly random mental image that is automatically generated in association with your current experience. Examples from perception might include driving without really concentrating on your surroundings and things going by in your peripheral vision. It would seem like a waste of brain processing power to require everything to be interpreted in order to perceive it; that is *even* if such an interpretation is possible for everything we experience. Hence I would suggest that this interpretation stage is reserved for more consciously willed and attended to episodes of consciousness, where a clearer conception of what is represented would be very valuable and hence worth processing. But to develop this idea we may need to distinguish between sub-personal and personal levels of interpretation and this is what I will discuss next.

6.2.2 - Different kinds of interpretation, examination and labelling

To look at different forms of interpretation, consider a willed episode of imagery where there is a specific goal in my mind, such as the Rubiks Cube problem. Here we need to work out how many small squares faces there are on a Rubiks Cube (this example was first introduced in Chapter 1). It could be argued that under the already adopted Kosslyn Model the processes involved in working this out could go something like this:

1. To consciously identify the need to visualize a Rubiks Cube in order to solve the problem.
2. For the information of what a Rubiks Cube looks like to be retrieved from memory.
3. For this information to be sent to fill the visual buffer appropriately.
4. For this to be accessed appropriately so a visual image is generated.
5. To then place close attention to one face in particular to count off the number of small squares.
6. To count of, or check, the total number of sides by mentally rotating the cube
7. And finally to perform some simple arithmetic to work out the total number of small squares.

The net result being that you are fairly certain that a single face is composed of 3 sets of 3 rows of small squares (nine squares per face) and that there are six faces per cube, yielding the final result that there are 54 small squares in total on the sides of a Rubiks Cube (6x9 =

54). Assuming the description above is fairly uncontroversial and matches most people's experiences of this process, which steps could involve certain forms of 'interpretation'?

I would like to identify two possible steps that require two different kinds of processes that might be close to what Tye means by interpretation. The first and most relevant in this context, is in the initial instigation of the episode of mental imagery. This involves the conscious willing of the generation of an image based on memories of what a Rubiks Cube looks like. In this case it seems plausible that this send-request needs to be coded in some kind of internal representation of the kind Tye has in mind: "Generate an image that represents a Rubiks Cube." This would somehow activate the OPP stream to search out this information from the associative memory of what a Rubiks Cube's visual parameters are. This may also involve the SPP stream in some minimal way to delimit the spatial parameters of the image, as discussed in Chapter 4. Once this information is retrieved it can be sent to the visual buffer in order to fill the arrays there. It seems likely that this will somehow retain an appended interpretation ("This represents a Rubiks Cube") so that it remains linked to the original goal. This would also allow the eventual mental image to come with an understanding of what it is and what it is being generated for. Hence the image comes pre-interpreted in a way due to the fact that has been consciously willed in the first place.

On top of that I would say there is another importantly different way that this image then needs to be subsequently 'interpreted', or more accurately *examined*; for instance to check the number of small squares and count the number of faces on a cube. For ease of reference I would like to re-term the first kind of interpretation described above as '**image labelling**' and reserve the term '**image examination**' for this second kind of image-interpretation-process. This latter process would involve *shifting your attention* to the relevant parts of the image and examining it to retrieve the appropriate information relevant to the task at hand (e.g. counting off the small squares). Hence, from now on, I will only use the term 'interpretation' when referring to Tye's use of the term, or when I am just using it according to its general meaning. It may be worth noting here that an appropriate Attention Shifting mechanism is already suggested in the Kosslyn Model as depicted in Figure 4.1.

How does this help clarify the situation? Well for a start it allows mental imagery to occur without a consciously accessible *labelling* process. This could occur by spontaneous (non-willed) filling-in of the visual buffer with information that defies any simple language-like interpretation or labelling. For example the abstract shapes and amorphous blobs suggested earlier could just arise by the filling-in and accessing of the visual buffer in an appropriate way. Secondly it allows for a differentiation between labelling that you are conscious off and some that you aren't, and allows a story to be told of how one can become aware of it if one attends to it closely enough. To illustrate this consider the example below.

> <u>Grandpa's Funeral Example</u>
> By visually noticing the anniversary date in my diary I am reminded of my grandpa's funeral last year. Soon after I may be surprised to realise I am experiencing an associated memory image of the front row of the people at the church service, and on examining it, I can subsequently remember who was sitting there.

I would suggest that this apparently spontaneously generated mental image was linked to the funeral by a fairly automatic sub-personal process that associated the two. This is what caused the un-willed instigation of the image and led to the seemingly spontaneous generation of the memory image, presumably by a relatively automatic filling-in of the visual buffer. Then on becoming aware of the image I am able to attend to the image and make explicit, by thinking about it further, the associated links of the context of the image and understand why it occurred when it did. In other words I can now become conscious of the spatio-temporal context of the image and have conscious access to why it may have been labelled in my memory as: “This represents an episode from your Grandpa’s Funeral in Oxford last year”. Perhaps we can call the initial sub-personal form of labelling a ‘**Label-S’** and the equivalent conscious personal level one a ‘**Label-P’**. So the image was originally associated with the funeral because it was labelled-S in my memory. Once retrieved and the image generated, I could examine and interpret the image and make the label-P explicit by thinking: “this represents an image of an episode from my grandpa’s funeral”.

Its worth noting however that I am not suggesting that there are two different kinds of labels here, only that there is one label that can be utilised in two different ways: either it is used only in sub-personal processes; or it can also be available at the personal level to used in a conscious episode of reasoning. This seems to explain the phenomenology of the experience and seems to be a good way of explaining the sub-personal processes that must be occurring some-how based on the neurons firing in your brain. I will use the term ‘label-S&P’ when the label is being used at both levels at once. This also emphasises the point that it is the same label and that there are just two different ways it can feature in our mental economy: i.e. by it just being appended to the imagery process in sub-personal mechanisms, or by also being aware of it at a conscious personal level.

To illustrate this utility further, this approach perhaps also allows the appending of new labels to new perceptual encounters, such as the fractal poster mentioned earlier. At first you just experience the phenomenal shape and colour aspects of the percept based on how it fills the contents of your visual buffer. Then you learn a general term used to refer to this category of fractal poster experiences, which then becomes associated with objects of this kind in memory and language and is thereby applied in future encounters of the same sort.

Note this *doesn't presuppose* the necessary existence of the label in order for this experience to occur, which seems to me like a fairly natural explanation. It also allows a neat explanation of situations where you can't quite remember the name of something but it feels like its on-the-tip-of-the-tongue. In this case perhaps the associations that would normally transfer the label-S to a consciously accessible level of label-P, aren't quite strong enough to bring it to mind presently, but you might keep working on it sub-consciously until it pops into your mind later.

If this categorisation is at least plausible, then we now have at least four possible ways that a mental image can be generated and accessed. **Firstly,** it can just arise spontaneously by subpersonal processes filling the visual buffer, where no labelling of any kind is required but a phenomenal experience can never the less be generated (e.g. an unattended meandering daydream). **Secondly,** it can be generated in a willed episode where conscious effort generates a label-S&P (i.e. at both levels) and this triggers a controlled episode of imagery. These forms of labels can remain associated with the image throughout the episode, as in the Rubiks Cube Problem discussed above. **Thirdly**, it can be triggered sub-personally with a label-S in associative memory. This may seem spontaneous at first, but on further reflection the association can be made explicit (label-P) to our conscious awareness, as in the Grandpa's Funeral example above. Although the association may be so abstract or tenuous that you can't ever really make the full connection even though you have a distinct feeling there is a connection, as in a vague feeling of knowing or recognising something. And **fourthly,** once the image is generated it can be *examined* or interpreted so that useful information relevant to the current goal can be drawn from it. And note this may allow new labels-S&P to be associated with it, for example "This represents a fractal poster" and "This represents just one kind of fractal pattern amongst many". So this seems to cater for a wider range of imagery experiences and this seems to match general reports of our imagistic experiences more closely. But is this an improvement on Tye's approach?

I would argue it is, because it covers a wider range of potential imagistic and perceptual experiences, from consciously willed ones to seemingly spontaneous ones. I say *seemingly spontaneous* here because they are obviously physically caused in the brain for some reason or another, its only that they may seem surprising and not obviously associated with anything from our personal level conscious perspective. Also, perhaps ironically, I might suggest that my approach fits more closely to Tye's underlying structure of PANIC theory in that the contents of the visual buffer only need to be accessed in a certain way for un-interpreted, non-conceptual, abstract and intentional experiences to be enjoyed. That is not only does it not have to be conceptual but it can also be uninterpreted at the representational level and they can be poised so that beliefs and desires can be generated from them is appropriate: i.e. this need not happen in all cases. This approach may also be able to cope more easily with other seemingly more spontaneous and un-interpreted vision-like states such as dreams and hallucinations and this is certainly one way I think this approach can be applied and developed elsewhere.

Another possible benefit of this approach is that it allows us to interpret the same phenomenal experience in different ways, perhaps by labelling the same contents in the visual buffer in a different way. For instance consider this *Sailor Picture Example*: here there is a picture of a sailor on a ladder that leads to a boat at the docks. The picture is meant to be ambiguous as to whether the sailor is going up or down the ladder and hence if he is returning from a trip or just setting off on one. It seems to me *an equivalent visual image* of the same scenario could be labelled either way and different associated concepts and labels might then be deployed to interpret it. For instance: if he is imagined to be going down the ladder and away on a trip, you will label it as such and there may be associated feelings of sadness for the loved ones he is leaving behind; conversely, if he is imagined to be returning then the label will change and there may be more happy associations; and perhaps if you are just checking the image to see if he is wearing a hat it may not matter either way[119]. Hence the very same image can be examined and labelled in very contrasting ways, even though arguably the contents of the visual buffer and the apparent *spatial layout* of the imagistic phenomenal experience could be identical in all three cases. What makes the crucial difference to your attitude towards the image is how it is labelled and whether this needs to be made explicit to examine it for information relevant to the task.

6.2.3 – Visual imagery and spatio-temporal labelling

The utility of this approach in terms of STIm should be evident here as it also potentially allows us to have certain images additionally labelled according to their temporal context and spatial locations. For example, a mental image of your room painted bright yellow might be labelled as: a memory of how it was in the past (past factual); how it could have been in the past (past counterfactual); or how it might possibly be in the future (future possible). In all cases the image conjured up could be look identical, but the label associated with it can represent a different temporal context. This also allows a neat way to explain the well-known phenomena of confabulation where imagined events are mistaken for real remembered events. For example, the actual image remains the same but somehow due to a processing error the label-S is switched from imagined (imagined possible) to remembered (factual past), and hence an imagined scene could get treated as if it actually happened. It’s also possible that the change in label also involves changing other aspects of your phenomenal experience. For instance, you may think it feels different to consider the

[119] You may also be surprised at what you find in the image having not realised you had imagined the sailor wearing a hat (or even that it was a male sailor that you were imagining). See Currie and Ravenscroft (2002: 26) for further discussion of this feature.

same image as representing an event in the past, compared with if you imagine it in the future. Appending this associated raw feeling of 'pastness' or 'futureness' is not essential for my approach, but it is certainly not ruled out[120].

Now obviously in the above I have assumed the subject has some concept of time with which to understand what these different temporal labels mean. But it is debatable point whether some way of imagining or remembering different temporal situations might be a necessary pre-cursor for even conceiving of other times. It may be that loss of an Episodic Thinking System, of which STIm may be a part, leads to a loss of any extended temporal awareness, as with certain subjects who seem stuck in the present with no memory or future directed thinking capacities[121]. It may even be that the capacity to *imagine* and *understand* other temporal contexts may have evolved hand-in-hand and are really two inextricable features of our ability to think temporally in the first place. This is certainly a debate that I am interested in exploring, but one that would involve a lengthy analysis that would go beyond what I can do justice to here[122]. Hence in this thesis I will restrict myself to modern human minds and only attempt to explain how they work in the present. So within this restricted context, I can safely assume that in normal situations, we already have some conception of past and future from which these labels can gain the significance that I need. This gives an indication of how the temporal aspect could be supported in STIm and how it could form the basis for further investigations into its role in our mental economy and perhaps even in supporting our ability to conceive of time itself.

---

[120] See Starwson (1994) and Horgan & Tienson (2002) for a discussion of how even mental states usually classified as non-sensory might have their own phenomenology. This might entail that considering different temporal contexts have a different feel to them. My theory is consistent with taking this position, even though I personally might not want to be quite as pervasive about what I consider to be phenomenal mental states.

[121] Although this is a complex matter and see Tulving (2002) for an example of a subject K.C. who can visually imagine things but has no memory of the past, or any idea of what he may do in the future.

[122] See especially Suddendorf and Corballis (1997, 2007) and Suddendorf and Whiten (2007) on the evolution of our ability to represent other times based on a capacity known as 'Mental Time Travel'.

In terms of spatial labelling it seems that you can imagine the same scene in very different circumstances. For example, you can imagine your bedroom as it is now in the arctic or in the desert, and then imagine what consequences there may be following from that. Note here that there are two different forms of spatial imagining going on. There is the apparent 3D distribution of the objects in your room and will I refer to this as the **spatial layout** of the image. And there is the geographic location that your room is thought to be in, which is a feature that might not manifest itself in the actual image, but may change your attitude towards the image and I will refer to this as the **spatial location** of the image. Regarding the latter, note that deploying this feature may also rest on the assumption that we have the ability to deploy and understand different spatial contextual labels that might get associated with these images. That is, in normal human beings we already possess the ability to understand that there are different possible spatial locations and hence these labels referring to them can be appropriately understood. That is even though the capacity to imagine a different spatial location and the ability to know that other spatial contexts are possible, may also be closely linked together in terms of their phylogenetic development. This is in a similar way to the temporal context labels discussed above and again something I cannot go into in detail here, but only assume that this capacity is present already in modern humans so that it can then be deployed[123]. This may also depend more on the SPP stream and our ability to keep track of where we are locally and globally (and even in the context of the wider universe).

I will return to be more explicit about these spatial and temporal context labels in the next chapter when I will illustrate what I think is a useful way of annotating them. This will then allow the imagistic contents of the visual buffer to, not only represent a specific spatial layout of certain objects, but this can also now be thought of as occurring under

[123] Interestingly there is less discussion of this aspect in the Episodic Thinking literature and may represent one way that my theory can help expand on and contribute to this field.

different temporal contexts and/or spatial locations. Hence covering all the aspects of STIm that I think are necessary for exploring its nature and function.

6.3 – STIm and the Interpretation Capacity

At the start of the chapter I suggested that analysing this part of Tye's work would help us elaborate in the Interpretation Capacity mechanism introduced in Section 2.2.3 and depicted on the output side of Figures 4.3. This mechanism is envisaged to be the one that allows us to examine the visual image and extract any useful information it represents that may be pertinent to the task at hand. Examples of this are counting or checking the number of sides a cube has and counting the number of small squares per face of a Rubiks Cube. I have associated this with interpretation-as-image-examination, and distinguished it from the use of the term 'interpretation' used by Tye. I have also suggested that perhaps using the terms *image-labelling* and *image-examination* helps disambiguate these two possible uses of the word interpretation in this context. I would suggest that the Interpretation Capacity as a whole would need access to both these facets in order to work properly. That is, it will not only have to examine the image, but how it interprets what the image represents will depend not only on the things contained in its spatial layout, but also the spatio-temporal context it is labelled as having. It will also be involved in helping direct attention to the pertinent items in the image, so it may work closely with the Attention Shifting mechanism as suggested by the Kosslyn Model (see above and Figure 4.1). In other words it needs to access any information relevant to the general interpretation of what the image represents and be able to use this to guide further thoughts and help instigate any further imagistic episodes that may aid this process.

Unfortunately I cannot go into much more detail in terms of explaining how the Interpretation Capacity does all this here due to space constraints. Hence I will only suggest briefly that, since imagery appears to us in the same visual style as vision (i.e. in shapes and colours), then perhaps the capacities with which we examine and interpret our online perceptions, may also be available for this same job in imagery. That is, I would suggest that all the knowledge and concepts that we have previously learnt and are able to deploy in terms of perception (e.g. folk physics, object names, emotions), should also be available to us in interpreting the meaning of visual imagery, because it presents itself in the same visual style modality. To now demand an attempt to explain how *that all works* for both perception and imagery, would be tantamount to asking for a solution to one of the major puzzles in Cognitive Science, and indeed Artificial Intelligence. Naturally I cannot attempt to do that here[124], but it may give one strong reason why it is useful to imagine non-current possibilities in a similar modality to how we perceive them. This is because we can re-use and apply all the same kinds of mental capacities, as we would use in interpreting online perception in terms of what certain shapes and colours represent. All this without having to develop a whole new set of mental processes exclusively for use with imagery.

Of course re-using these interpretation mechanisms used in perception would also require making sure they are applied in a mock or offline way when applied to mental imagery. For instance by making sure any conclusions you draw from experiencing a visual image of ‘a tiger in front of you’, are not then treated as if there really is a tiger in front of you. This could be realised by making sure all these interpretation processes maintain a mock label of some sort, as would the visual image. Hence this re-use of perception-based concepts and reasoning processes, in order to examine and interpret imagined episodes, seems like a feature that again supports the parsimony of an imagination-as-simulation

[124] This mirrors a similar claim by Stich and Nichols (2003: 30), that any further demands to explain which previously held beliefs get updated or not based on a new belief, would also be unreasonable, because this would be tantamount to solving the Frame Problem ‘which has bedevilled cognitive scientists for decades.’

approach. So while I have not gone into much detail about the Interpretation Capacity here, I think there is no real mystery surrounding why it should occur. This is because we can all probably accept that the same thing must occur in online vision and that this same capacity seems like the right kind of thing to apply to visual images, since they are both directed at similar features, like the shapes and colours of objects laid out before us.

So in summary, the main aim if this sub-section was to show how Tye's interpretation-based approach might be seen as being too constrictive and inflexible, and hence to provide a platform to develop my own alternative label-based approach that seems to me to explain more of the data and to be more flexible. It may be noted that I have offered no empirical evidence to support this label-based approach and it is merely offered as some theorising of what I think must be going on somehow. However there is some support for this because we may just be appealing to a more propositional-style approach to mental representation in order to develop this labelling aspect, and this is already generally fairly well accepted in the literature. Thus all I am trying to do here is to suggest a way a more propositional-style kind of labelling process might interact with the more depictive element that is the basis for generating the detailed spatial layout of phenomenal imagery. And although this labelling seems to me to *not be essential* for imagery to occur, it seems likely that it will be required for any considered episode of thinking that utilises imagery[125]. It certainly seems that my labelling-based approach could be used as a basis for further empirical work or as a basis for interpreting already existant evidence in this area.

[125] The use of imagery in thinking and analysing how reliant we are on language to reason, is something I would like to analyse further in the future. Imagistic thinking certainly opens the possibility that other animals might utilise it without requiring them to be able talk. This is why I think this approach may be useful for developing ideas in the Episodic Thinking literature. See also: Bermudez's (2003) book entitled 'Thinking without words' and a review of it by Fodor (2003).

I have now dealt in Chapter 6 with the two aspects of Tye's work that I think I needed to, before moving on to develop my own positive account in Chapter 7. Specifically in the above, I have rejected his Reductive Representationalism PANIC theory and developed and expanded on his interpretation-based approaches to of *perception and imagery*. In doing so I have suggested why I think a form of non-Reductive Representationalism is to be favoured and why my labelling-based approach to mental imagery has significant benefits. Further to that, and more generally, it is worth noting that I think have now looked at all the individual mechanisms depicted in Figure 4.3, in as much detail as I think is appropriate for each one; and certainly this is in as much detail as I have time for here. So perhaps I can now try and draw all these strings of evidence and reasoning together in the next concluding chapter, in order to develop my full theory of STIm. Given what has gone before, this will obviously be in terms of imagination-as-simulation, which in turn will be based on the Kosslyn Model, and which will appeal to a Generative and a non-Reductive Representational Approach.

7 – STIm: the Full Theory

A Representational and Generative Approach
based on the Kosslyn Model as an imagination-as-simulation process

> Let me add, finally, that the [local] supervenience thesis enjoys further support from extensive research in neuroimaging. Every aspect of experience, from illusory contours to motion illusions, to phantom limbs to diffuse pain, can be correlated with some neuronal response…. Neuroscience shows every promise of being deliciously complete… Many philosophers would balk at this claim. Neuroscience is incomplete in one crucial respect: it can't explain why all these neural correlates feel the way they do. It can't even explain why they feel like anything at all. Perhaps this is the missing element that makes Noe feel like he has to go outside the head. If so it is a fool's errand. For this deep epistemic problem of how physical states can be phenomenal experiences is in no way ameliorated by broadening the supervenience base. Just as it's hard to understand why brain states feel a certain way, it's hard to understand why brain states together with bits of external environment feel a certain way.' Prinz (2006 – on a review of Noe 2006)

> 'What I cannot create, I do not understand.'
> – the inscription left on Richard Feynman's blackboard after he died

In the last chapter we looked at two different aspects of Tye's (1991, 1995) work that concerned the nature of perception and visual imagery. These were his PANIC theory of phenomenal consciousness and his interpretation-based approach to imagery. There I rejected his overlying PANIC theory and I developed my own label-based theory that expanded on Tye's interpretation-based approach. I had also suggested there that these were the last two features that I needed to address before developing my own full theory of visual imagery. Hence in this chapter I will now introduce and defend my preferred imagination-as-simulation approach to STIm under a Generative Approach to perception. The majority of this chapter will consist of a fairly positive exposition of my Conjunctivist Position, which suggests that, contrary to what Disjunctivists might claim, perception and imagery do have a significant common overlap in the kind of things they are. This claim will be explained further below, where I will also draw together most of the threads of the arguments from previous chapters and develop them under a non-Reductive Representationalist framework.

A non-Reductive Representationalist position suggests that while perception is essentially Representational (as opposed to Relational), the phenomenal character of experience cannot be reduced merely to its representational contents. Here I will suggest that some further *generative* step is required where the experienced visual field is produced based on the neuronal representations in the visual buffer. I will argue that these same processes might also be available for the generation of phenomenal imagery based on accessing the offline contents of the visual buffer. Hence I will utilise and adopt the claims put forward by the Kosslyn Model, which suggests that both vision and visual imagery utilise the visual buffer and these are in turn based on an underlying hybrid depictive representational style. The main aim of this chapter is therefore to look at phenomenal imagery from either side of the symbol to phenomenal gap, by explaining how these underlying representations in the brain may eventually underpin phenomenal imagery. The conclusions from this will then be merged with my labelling-based approach for setting the spatio-temporal context of a visual image and this will yield a final full account of STIm by the end of the chapter.

So specifically in this chapter, I will **firstly** briefly repeat what I think I have established so far and hence what I will assume as my platform for the following positive arguments. This will also act as my main summary of the thesis before my main conclusions later in the chapter. **Secondly**, I will re-iterate and develop in more detail what I see as the relationship between the visual buffer (cells) and the visual field (pixels). This can then be applied to describe the relationship between the visual buffer and phenomenal imagery. **Thirdly,** I will suggest that this can then be used to explore the relationship between STIm and phenomenal consciousness under a Generative Approach to perception. I will explore some options for what this generative process might be based on, but will not attempt to explain the mechanism itself. And then **finally,** I will put forward my main proposal for STIm, which will include how a generic visual image becomes associated and labelled with different spatio-temporal contexts. At this point my theory will have explained all of the features I have set out to explore in the thesis and hence I can then give my main conclusion. To that end I will start this chapter by summarising the claims I have made so far in the thesis.

7.1 – Thesis summary - the claims so far

The key claims of the thesis so far have been as follows. In Chapter 2 I introduced some examples of different simulation mechanisms and went into detail about the kinds of features we should look for in an offline simulation process. I also discussed a potentially useful way to distinguish between emulation and simulation mechanisms. This gave us an insight into the current debate in this field and illustrated some ways that simulation mechanisms are being utilised in certain recent theories to explain mental features related to STIm: e.g. motor imagery and visual emulators (Grush 2004) and processes that might underlie our capacity to *deliberate* (Hurley 2008). This acted as a way to illustrate what I mean by offline simulation process, which in turn provided the template with which to investigate whether the Kosslyn Model can be interpreted as an offline simulation process in the next two chapters.

In Chapter 3 I introduced the hybrid depictive representational approach favoured by the Kosslyn Model and defended it against a propositions only approach (c.f. Pylyshyn 2003). This allowed me to suggest the following way of annotating the hybrid contents of array cells in the visual buffer:

Array Cell (X, Y) = [colour (RGB: P, Q, R), depth Z, surface orientation B, edge D, etc…]

.

This involved a mixture of depictive and proposition-like elements and hence yielded a hybrid representational approach. The depictive element stems from the topographic way that these images are represented in the visual buffer and the proposition-like element comes from the information stored in each of the cells in the array.

In Chapter 4 I took a close look at the Kosslyn Model and defended it against some putative counter examples. These threatened to show that detailed depictive imagery could occur without the visual buffer, which would count against this kind of approach. I also looked at deficits in vision and visual imagery in terms of Neglect and offered an explanation of how information in the ventral (OPP, what) stream and dorsal (SPP, where/how) stream[126] could interact and be affected together or separately. This suggested one way that the *spatial layout* of a visual image could be delimited by integrating associative memories of an object's shapes and colours with a (possibly minimalistic default) spatial layout set by the SPP stream. We then looked at a couple of options for potential ways that the visual buffer could be filled in internally in an online or offline (mock) way. One of these offline options included some tentative evidence for a visual emulator mechanism and the other was a version of the Kosslyn Model's search priming process. The combination of the features discussed in Chapter 4, allowed me to suggest the Kosslyn Model could indeed usefully be interpreted as an offline simulation process. This

[126] Recall from Chapter 4: OPP stands for 'Objects Properties Processing' and controls the colours and shapes that objects have in imagery and runs forward (ventrally) from the visual cortex through the temporal lobes; and SPP stands for 'Spatial Properties Processing' and controls the spatial orientation of the objects in imagery and runs up through the back (dorsal) part of the brain from the visual cortex to the parietal cortex. See also Appendix B.

also suggested the following separation of ways the visual buffer could be filled-in and labelled:

Perceived: (online, external) {Array Cell (X, Y): colour (RGB: P, Q, R), depth Z, etc}

Imagined: (offline, internal) {Array Cell (X, Y): mock, colour (RGB: P, Q, R), depth Z, etc}

At this point I concluded that the Kosslyn Model could indeed be usefully interpreted as an offline simulation process and seemed to be a good theory to adopt for the remainder of the thesis. This is because it is fairly well corroborated and the same mechanism (the visual buffer) is used in vision and visual imagery, but in imagery it is only used in an offline, or mock, way. This also allowed me to account for ways that the image could be maintained and transformed and how this could account for certain other aspects of the phenomenology of imagery: e.g. scanning and rotation tasks and that it is normally less vivid and quickly fades away unless a concerted effort is made to maintain it. Hence yielding a good combined theory of the sub-personal processes underpinning imagery, which could be used as a platform for exploring the more phenomenal aspects of visual imagery in what followed.

To begin Part II of the thesis I explained that it might help to understand what is happening in the offline simulation of visual imagery, if we understood what might be going on in the online perceptual case. To that end I identified the most important distinction to this analysis as being that between Selective and Generative Approaches and indicated that I would reject the former in favour of the latter. This would allow me to suggest that the brain generates phenomenal experiences in a fairly similar way in both perception and imagery and hence leads me to adopt a Conjunctivist Position. To that end I explored and rejected what I have interpreted as two different kinds of Selective Approaches to perception: Naïve Realism and Tye's PANIC Theory. In Chapter 5 I introduced and explained why Naïve Realism, which I took as a paradigmatic example of a Selective Approach and Disjunctivist Position, may seem at first like an attractive theory to adopt, but I also explained why I don't share some of the usual motivations that are given for defending it. I then offered the Time-spread Argument as my main reason for rejecting it and briefly indicated that this also might lead me to reject all Selective Approaches to perception.

In Chapter 6 I criticised Tye's (1995, 2000) PANIC theory for being overly representationally reductive and not really offering an explanation for how PANIC states support phenomenal consciousness. However at the same time I acknowledged that much of his work describing the mechanisms underlying vision and visual imagery seem to be compatible with more recent work on the Kosslyn Model and hence at least this part of his theory remains highly compatible with my theory. However I have interpreted Tye's full PANIC theory as being most naturally classed as a Selective Approach and hence I have rejected this overlying explanation of phenomenal character by appealing once again to the Time-spread Argument. Finally I have given arguments defending my replacement of Tye's (1991) interpretation-based approach to mental imagery, with my own 'labelling' approach, which I felt explained a wider range of our imagistic phenomenology in a much more intuitive way. This also provided me with a way to explain how the spatial layout of a visual image could also be associated with different temporal contexts and/or different spatial locations, by simply being labelled as such. Hence allowing it to be thought of in different spatio-temporal context and this provided the final pieces I felt I needed to give a full account of STIm in this chapter.

The main conclusions from Part II so far are therefore as follows. Due to the Time-spread Argument it seems implausible to me that phenomenal character isn't locally supervenient and this allows me to now assume some form of Representational and Generative Approach to perception. I refer to my theory as non-Reductively Representational because it accepts that phenomenal character is generally representational, but denies that it can be always reduced to representational contents and hence something else is needed to explain the phenomenal aspect. And while I will not be specific as to how this extra generation step works, my theory will be broadly compatible with those that accept that the brain generates phenomenal character in some way and hence accepts some form of a local supervenience constraint (e.g. sense-data and qualia theories). Reaching these conclusions also allows me to suggest here that if the brain generates phenomenal character in the case of perception then perhaps this same feature can be used to generate phenomenal imagery. How this can be developed in detail is the main focus of what follows and I will start by looking at the structural relationship between the symbolic contents of the visual buffer and the our phenomenal experience of the visual field next. This will leave the generation step as the missing link to cross the gap between the two sides and this will be discussed in Section 7.3.

7.2 – The relationship between the visual buffer and the visual field

At the very start of the thesis in Section 1.3, I introduced some evidence which suggests that damage to areas of the visual buffer that are farther apart, leads to scotomas (or blindspots) in the visual field that are also farther apart (c.f. Kosslyn et al 2006: 15 and also see my Figure 1.4). Hence there seems to be some correlation between damage to areas in the visual buffer with a related loss in the visual field. This is to be expected if under the Kosslyn Model we are suggesting that the topographically organised areas of the visual buffer correspond to the way light is detected at the eyes retinotopically and this in turn is related to where the light has come from within the part of the world that corresponds to the visual field. Hence it is fairly well predicted and supported that under this model, topographic areas in the visual buffer correspond to areas of the visual field. And recall that central areas of the visual buffer seem to have more cells devoted to them in order to support more detailed experiences at the centre of your attention (c.f. Section 3.1.2).

Further evidence for this includes correlations between which parts of the visual buffer are accessed and utilised in relevant episodes of perception and imagery: e.g. in rotation and scanning tasks. I have detailed a few experiments that support this interpretation[127] in Chapters 3 and 4, but will explain one further one here to make this even more explicit. This is where the size of the perceived object, or the size of an imagined object, correlates with the amount of visual buffer activated (c.f. Kosslyn 1978). This means that objects that occupy more of your visual field also require a larger area of visual buffer to be activated. The visual size of an object can be best described as the visual angle it takes up in the visual field. A large object far away will not occupy much of your visual field and hence occupy a small visual angle, but when it is close up it will occupy a large visual angle. So the evidence above is explained as being due to the greater the visual angle that the object occupies, then there will be more cells in the visual array required in order to process it. It seems that this feature is duplicated in imagery tasks, where imagining objects that would take up more of the visual field, also activate more of the visual buffer. And further it seems that in some cases, a drop in the size of the available visual buffer, perhaps due to an operation, can also correlate with a corresponding drop in the maximum size of the image that one can imagine. This also corresponds to a drop in the size of the actual visual field available at the personal level. Hence we seem to have another fairly good case for correlating areas of the visual buffer with areas of the visual field.

[127] See also Kosslyn et al (2006: 108-130) for a more detailed analysis.

In addition to this evidence there is also some interesting new research emerging that builds on these assumptions. In recent studies, pattern recognition programs are being trained to predict what a subject was looking at by analysing a brain scan taken at the time. This is done by the program interpreting fMRI brain scans of the topographically organised areas in the visual cortex and from that attempting to reconstruct the shape of the original external visual stimulus. At first the program is trained by matching a copy of the shape of the original perceived stimulus with a scan of the relevant brain area and this allows it to form a method and generate algorithms for predicting future stimuli. Then a new brain scan of a novel stimulus is given to the program and it can fairly convincingly reconstruct the shape of the original external stimulus. For example Kamitani et al (2008) showed subjects the letters in the word ‘neuron’ in sequence and the program reconstructed these as depicted below:

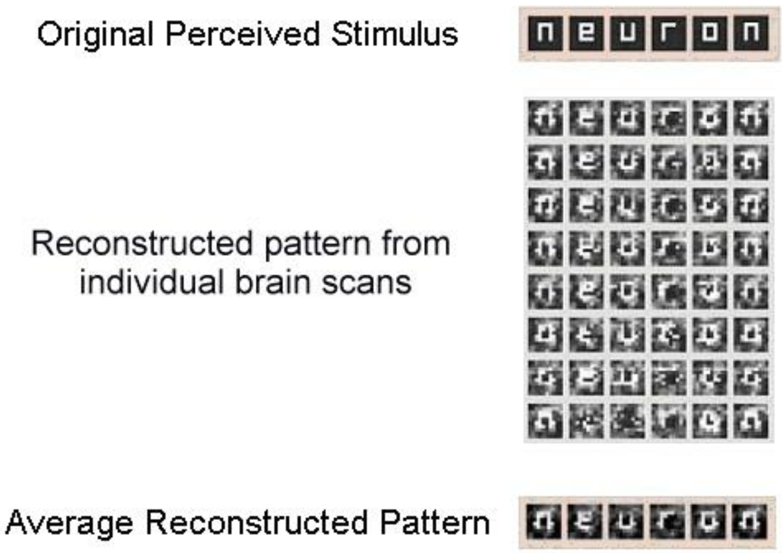


Figure 7.1- showing the perceived stimulus (above) and the averaged reconstructed pattern (below) worked out from brain scans of relevant topographically organised areas of the brain after initial training – as adapted from Kamitani et al (2008)

These results and the assumptions they employ seem to re-enforce quite positively the relation between areas of the visual field and topographically organised areas of the visual cortex. This has lead the researchers conducting these experiments to speculate that the same principle could be applied to decode *offline imagined* states after the program had been trained using the *online* perceptual situation first, as this quote might suggest:

> 'More interesting are attempts to reconstruct subjective states that are elicited without sensory stimulation, such as visual imagery, illusions, and dreams. Several studies have suggested that these subjective percepts occur in the early visual cortex (Kosslyn et al., 1995), consistent with the retinotopy map… Of particular interest is to examine if such subjective percepts share the same representation as stimulus-evoked percepts… One could address this issue by attempting to reconstruct a subjective state using a reconstruction model trained with physical stimuli.' Kamitani et al (2008 – other references removed from quote)

I interpret this as meaning that it may be possible, once the pattern recognition program is trained to interpret the brain scans from online sensory stimulus, that this could then be used to reconstruct what someone might be visually imagining. For instance, instead of seeing the letter 'N', the subject could focus on visualizing it and the program could then attempt to reconstruct what one is imagining by analysing the relevant brain scan. The idea of this possibility is repeated by other researchers like Kay et al (2008) who suggest that such a device could potentially decode what one was imagining or dreaming[128]. If this process worked then it would be fairly good confirmation of the Kosslyn Model and indicates how it is being used as a basis for further pioneering work. Hence this is an exciting time to be in this field and further experiments might soon be able to confirm the ideas presented here one way or the other. The reason I introduce them here is as further recent evidence for the claim that cells in the visual buffer and pixels in the visual field are closely related.

---

[128] It is interesting to note that the 'Big Brother' mindreading implications of such a device were quickly associated with these studies in the media at the time (see also the New Scientist article on this in the 20th December 2008 issue).

At this point it is perhaps worth recalling that at the start of Chapter 2, I introduced a warning from Hurley (2008) that suggested we should be wary of 'interlevel isomorphism assumptions' between personal and sub-personal levels of description. In this context this might mean we should be careful not to automatically assume that neuronal structures at the sub-personal level in the visual buffer, are correlated to the structure of our phenomenal visual experience at the personal level. With this warning in mind, and given the above, it may be that in this case some kind of interlevel structural isomorphism is justified. That is, it seems that we have good evidence that links topographically organised areas of the visual buffer with areas of the visual field. Perhaps then this is no longer counts as a mere assumption, but is instead a well supported hypothesis. Hence, whilst I feel I have heeded Hurley's warning, the evidence does seem to justify accepting a certain amount of interlevel structural isomorphism in the case of vision and visual imagery at least.

In Chapter 1 talked about referring to small areas in the visual field as 'pixels' in an analogous way to pixels on a TV screen. I also made the corresponding warning there that naively using this term might be misleading, because it might be taken to indicate I think the visual field is *actually* split up in this way. However this is not something I need to, or want to, argue for. I can use the term 'pixel' here to simply refer to a small, and perhaps only-just-discernable, area of the visual field, without also arguing that the visual field is actually split up in this way as a TV screen might be. With these warnings in mind then, it therefore seems possible to draw some correlations between cells in the visual buffer array and 'pixels' in the visual field. But again I would highlight the fact that I am not suggesting there is a one-to-one relationship, only that activity (or lack of it) in certain cells of the visual buffer does seem to correlate with phenomenal experience (or lack of it) in certain areas or pixels of the visual field. It may be that several cells are needed to support one pixel, or that there is a more complicated interaction between nearby cells and upstream processes, that eventually affects what one experiences in a certain area. I do not need to take a more specific stance on this as long as there seems to be some close correlative link between areas of cells and pixels, then I think this is enough to support what I want to talk about next.

So given the above perhaps I can tentatively suggest the following example to illustrate these points. Consider a case where you see, or are asked to imagine, a red rectangle on the left and a blue square on the right. And consider how this might change if you were asked to see, or imagine, a blue rectangle on the right and a red square on the left. Perhaps we could annotate the simplified contents of the visual buffer in the diagram below, where I have only specified the colour detected in each cell. That is, all other content apart from those representing colour have been removed for simplicities sake (and see Figure 3.3 for my first use of a diagram like this):

1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	**8**	**8**	**8**	**8**	1	**5**	**5**	**5**	1
1	**8**	**8**	**8**	**8**	1	**5**	**5**	**5**	1
1	**8**	**8**	**8**	**8**	1	**5**	**5**	**5**	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1

1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	*5*	*5*	*5*	*5*	1	*8*	*8*	*8*	1
1	*5*	*5*	*5*	*5*	1	*8*	*8*	*8*	1
1	*5*	*5*	*5*	*5*	1	*8*	*8*	*8*	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1

Figure 7.2 – Visual Buffer Matrix Annotation for:
above –red rectangle on left and blue square on right (seen);
and **below** –blue rectangle on left and red square on right (imagined)

In the situation above, cells in the visual buffer are filled in with values that represent the retinotopically distributed light wavelengths detected at the eye. These will eventually get accessed by upstream mechanisms and experienced as shapes and colours in corresponding parts of the visual field. Of course this is vastly oversimplified to make the point and other information in each cell has been left out to make things simple. For example information like the depth (or distance to object) values can be assumed constant. I have tried to illustrate the two cases slightly differently, by filling in the top set in with numbers in bold, and the bottom set in with numbers in italics. This is intended to indicate the potential difference between how real *online* inputs and *offline* mock inputs could be represented in the visual buffer. For example: the top matrix might represent a visual episode and the bottom matrix might represent an imagined episode.

Again I stress this is a very simplified schematic and just serves to illustrate the point that similar cells in the visual buffer might be filled-in in each case in slightly different ways to keep them separate during further processing. This way of looking at it is consistent with the evidence that the same areas are used in vision and visual imagery and that they compete for resources in both processes. And importantly this would still count as an offline simulation of the online visual process because the visual buffer is being used in both cases. But note that this interrelation of offline and online entries could be very complex and it is possible that they can co-exist in the same cell and be distinguished by the presence or absence of a mock label. This inter-cell and intra-cell co-habitation of the mock and real entries in the visual buffer, and how they might compete for our limited attention, is certainly something I would like to analyse further in the future, but unfortunately is all I can say about it here[129].

In terms of now describing how the situation above relates to the visual field, I offer the illustration below, which is an adapted version of Figure 1.4, which was the scotoma diagram from Chapter 1. In the below I have *coloured-in* the cells of the visual buffer rather than *filled-them-in-with-numerical-values* for ease of illustration. But if you prefer, it is easy to transpose the numbers from the above example (e.g. the 5's and 8's) onto the visual buffer on the left hand side of the diagram:

[129] It may be worth noting however that the visual buffer is composed of millions of neurons so this kind of complexity could easily sustained. And this might also help prevent one from thinking of the cells as literally bucket like slots in a matrix where only one thing can fit in it. The reality of the situation may be far more complex than that.

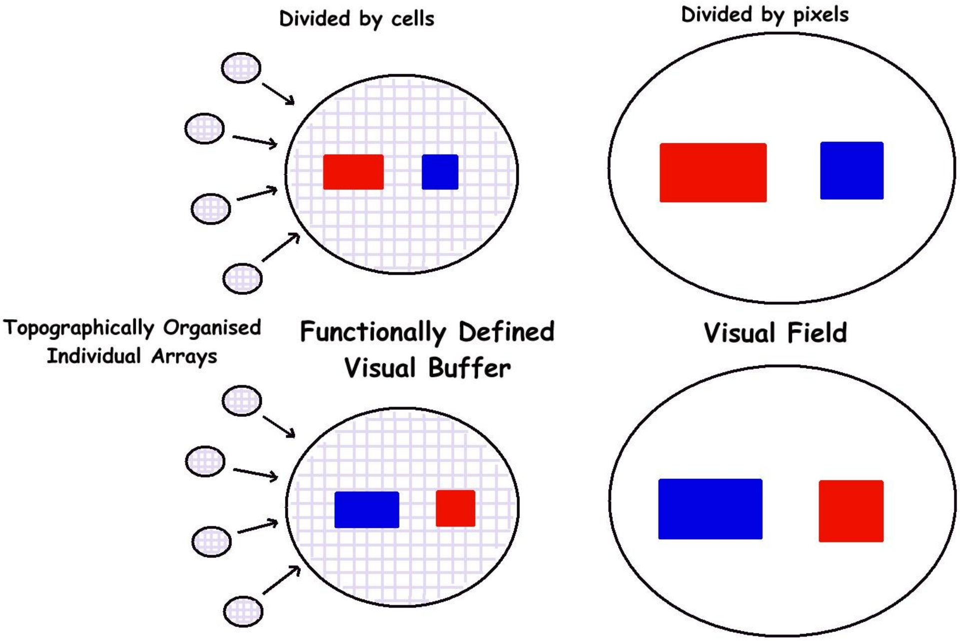


Figure. 7.3 – Visual buffer cell to visual field pixel relation for the situation described in the text and the situation described in Fig 7.2 above

If the above is plausible, it perhaps offers one way that we can think of the cells in the visual buffer corresponding with perceived objects in the visual field. And given what has gone before it seems reasonable to suggest that this can also be related to how the spatial layout of a visual image can be controlled and delimited with certain crucial differences which I will say more about below.

It seems to me that a similar set of diagrams could be made for objects that are uniform in colour but vary in depth, where the Z-coordinate would vary in the visual buffer in proportion to distance to object in the visual field. Although the description above has necessarily been very simplified, I suggest that this is more to do with the need to describe it in a mainly a text constrained fashion. To avoid these constraints I see no reason why this couldn't be developed and demonstrated in a more complicated and involved way, perhaps on a graphics package on a computer. Perhaps this could be done by representing more than one cell content at once and displaying many more pixels in a much finer grid.

Hence the above can be seen as an exposition that sets the style of explanation rather than any real attempt to indicate the complexity of the actual situation. So now we have a correlation between the layout of the visual buffer and our phenomenal experience, these leaves an explanatory gap between how one is actually produced from the other. I take this to require some kind of generation step to bridge the symbol to phenomenal gap and exploring the extent of what we can reasonably say about this at present will be the focus of the next section.

7.3 – Visual imagery, phenomenal consciousness and the generative process

On the assumption that the mechanisms underlying visual perception can be characterised in a cell to pixel correlation under the Kosslyn Model as described above, what can we say about how this relates to Generative Approaches to perception? What I would want to say here is that what ever process that you think generates phenomenal character in the online perceptual case, can now perhaps be commandeered to generate the phenomenal character in the imagistic case. This would indicate that imagery and perception are produced in fairly similar ways, with certain minor differences in how they are caused and experienced, and these features will be discussed below. The conjunction of these different features under the same process is illustrated in the following diagram, where the crucial step is that they all go through a *generation process*, which converts them from merely symbolic information in the visual buffer to phenomenal experiences of shapes and objects:

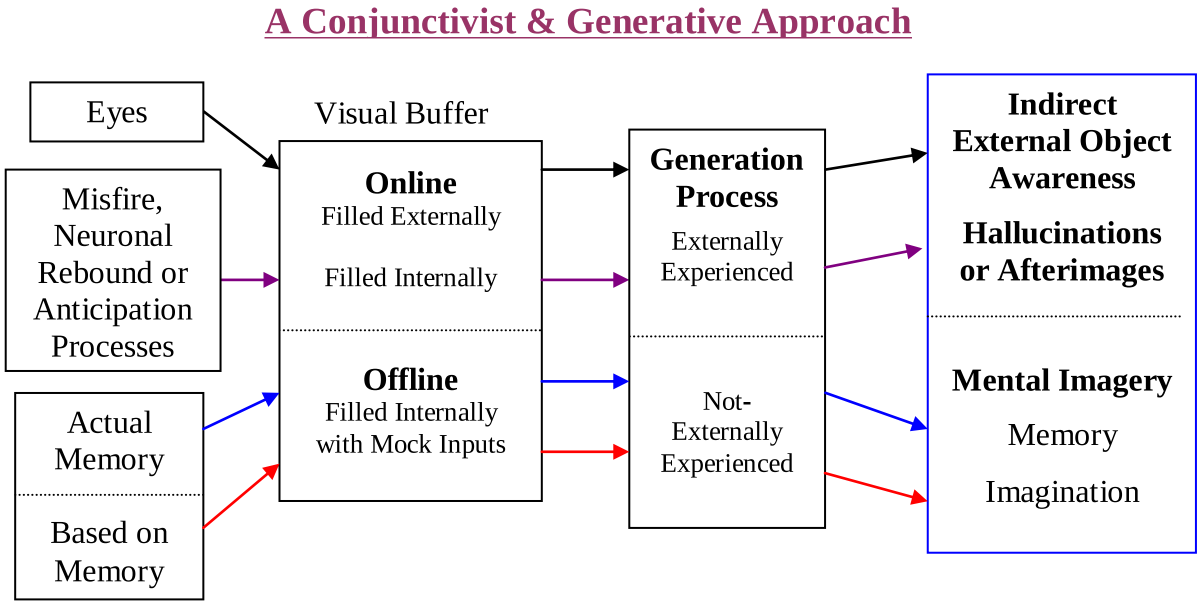


Figure 7.4 – Schematic of a Conjunctivist & Generative Approach to vision and visual imagery: assuming an imagination-as-simulation approach based on the Kosslyn Model.

Note that I have included on the left some different external and internal stimuli that might lead to filling-in the visual buffer in different ways. These have been linked by matching the colours of the arrows in the processing flows to the respective eventual mental states that they may lead to (on the right). Since I have already attempted to explain sub-personal mechanisms that fill-in the visual buffer up the point of the generation process, I will focus mainly on the right hand side of the above diagram in what follows.

It is also worth noting here that the above diagram explicitly illustrates why my theory can be clearly classed as a form of Conjunctivism. This is because it conjoins or unifies: perception *and* illusions *and* hallucinations *and* imagery, as all being *locally generated visual style phenomenal representational mental states*. This means they are all: visual style[130] phenomenal states in that they are all experienced in apparent shapes and colours

[130] Recall: I first used this term in Chapter 1 to refer to any mental state that is experienced in apparent or real shapes and colours. This is as opposed to vision-like states, which includes visual hallucinations and visual

(this includes visual states themselves); and that they represent things that one is seeing or might see; and they are all generated locally by the brain; and hence they are ontologically *similar things in at least this respect*. That is they are relatively similar things when compared to non-phenomenal mental states and even when compared to different types of phenomenal mental states in other sense modalities: e.g. hearing and taste. This is because these other states need not share the same underlying topographic processing style and extensively overlapping neurology (especially not the visual buffer) and will not lead to the generation of the same style of phenomenal experience (i.e. they wont be experienced in apparent shapes and colours in normal subjects). This is what puts these vision and vision-like states together as a group and why we can say they share a significantly higher common factor when compared to other types of mental states.

It may also be worth emphasising that I think this claim is crucially different from claiming that these vision and vision-like states are all ‘the same *kind* of thing’, since this might then require an analysis of what one means by ‘kind’ and what might count as a *natural* or *fundamental* kind in this respect. It is possible that vision, illusion, hallucination and visual imagery all share a significant common factor and yet could be classed as different natural or fundamental kinds of things at some higher level of description. This may depend on your definition of what counts as a natural or a fundamental kind[131]. The claim here is merely that they share a significant common factor and are therefore *relatively similar* (but not identical) kinds of things. This similarity is especially emphasised when compared to

---

imagery, and these are like visual states in that they have the same visual style, but have different causal histories.

[131] For a discussion of natural kinds see Hacking (1991) and also see Pautz (2007) for a criticism of appeals to *fundamental kinds* from Naïve Realists like Martin (2004,2006). For instance Pautz (2007: 528) claims: ‘I do not understand the notion of a fundamental kind. Neither Wiggins [1980] nor Martin gives any examples… And since it is very difficult to justify the claim that there are such things as fundamental kinds at all, it is very difficult to justify Naive Realism as Martin formulates it.’ Hence I prefer not to use the term ‘kind’ in any technical sense in what follows but more as a general synonym for ‘type’: for example, as in ‘a similar type or kind of thing’.

the large differences between these states that might be claimed by a Disjunctivist, as discussed in the last section and illustrated in Figure 5.1 for instance[132]. So while the Conjunctivist accepts there are some minor differences between these states, we would claim they are all still relatively quite similar, and share a significant common factor in that they are all *locally generated visual style phenomenal representational mental states*.

In addition to that the Conjunctivist can still claim this similarity in the types of mental states they are, is still compatible with the fact that the way that they are caused and experienced may be very different. I think I have explained in detail the differences in how they are caused in the first part of the thesis, and have developed the annotations given above to reflect this. Explaining the differences in how they are then experienced is the topic of Section 7.3.2, but first I will look in the next section at what I think I can sensibly say about the *generation process*.

7.3.1 – The generation process and phenomenal consciousness

So in terms of the generation process itself, I will not attempt any explanation of a mechanism by which consciousness is actually generated. Nor will I try to describe explicitly how the symbol to phenomenal gap is crossed, since I think this remains a fascinating mystery at this point in our understanding. I also think it is far too early in our understanding of consciousness to say anything very strong about whether on the limit we will ever be able to understand how consciousness can be a physical process. My opinion is that modesty and honesty at this point is the correct emphasis to have[133]. As a result, all I

---

[132] There I suggested that under Naïve Realism that perhaps: online visual states could be classed as widely supervenient and put one in touch directly with the object's properties of shape and colour; but vision-like states like imagery could be locally supervenient and their apparent shapes and colours would be generated by the brain.

[133] And I agree with Shallice (1997) that it would be a-historic to assume the problem of consciousness is unsolvable, unlike many other at-the-time seemingly mysterious physical phenomena e.g. planetary motion and fire.

will say is that based on accessing the information in the visual buffer by other upstream (or same level) brain areas, *somehow* phenomenal character is generated, and this is what we experience in phenomenal consciousness. This means that what is represented in the visual buffer is necessary, but probably not sufficient, for full phenomenal consciousness. However, as discussed in the above, the contents of the visual buffer probably tightly control the parameters of what is perceived, even if they don't control the final explicit phenomenal *way* that it is perceived. That is they provide the basic symbolic representation of what is out there in the world, but this can still be interpreted slightly differently by processing quirks further along the visual process.

As an illustration of this flexibility: this **allows** for something like the Muller Lyer Illusion[134] to be potentially explained by processes further upstream. For instance this may be because it occurs due to certain processing shortcuts to do with considering the context of the attached arrowheads, which means the raw data is interpreted as lines of different lengths. This may be despite the fact that the raw information in the visual buffer might represent them as being the equivalent of equal length lines (e.g. equal number of cells filled). Note this explanation is merely *allowed* for here, rather than suggested. The Muller Lyer Illusion could still be caused by processing quirks within the visual buffer itself or further downstream and I take no strong stance on this issue. What we could say though, is that these illusory cases are probably rare and our experience will usually match up fairly closely with the information accessed from the visual buffer. Which in turn will also match quite closely with the external world itself. That is, these kinds of illusions are the exception rather than the norm and our perceptual experience is most often veridical (even if not as direct as some might hope for once a Generative Approach is adopted).

[134] The Muller-Lyer illusion consists of two equal length horizontal straight lines with arrow-heads on the end. The arrow-heads point inwards on one line and point outwards on the other. This gives the illusion that the lines are different lengths when on measuring them they are clearly equal.

As mentioned above although I will not attempt an explanation of how the signal to phenomenal gap is actually crossed, I will make a *relatively* weak assumption that some physical story will eventually be told about this generation process. That is even though I accept that, in the limit, some very modest form of dualism may be necessary and we just don't know if this will be required yet; though and I am optimistic that it wont be[135]. I say 'modest dualism' because I think it will still be tied very closely to the workings of individual brains and will invoke as little as possible extra in terms of non-physical entities: i.e. just enough to explain how brain states can generate phenomenal consciousness and no more. However, as mentioned at the end of Chapter 5, at the moment I take appeals to dualism to be premature and are equivalent to appeals of the last resort, since they seem to just open more questions than they answer at present and hence offer no real benefits whilst incurring the invocation of strange non-material entities. The important point I want to stress here is that I take it that the same generation process could still occur in vision and in visual imagery, whether it is a purely physical process or not. Hence I don't see this assumption as very crucial to my Conjunctivist Position and the above is merely an indication of my position on materialism for the benefit of the reader. So while I have not pretended to offer a complete theory of consciousness, I hope to have given a fairly comprehensive account of the implications of this to a simulationist approach irrespective of whether materialism is eventually found to be true or not.

---

[135] I think this makes me a Type-C Materialist according to Chalmers (2003: 119) or 'physicalist hopeful' according to Tye (2000: 22), but a detailed categorisation of my position is not essential here. For the record I have some sympathy with a position called 'Subjective Physicalism' put forward by Howell (2008). And see Block (2003) and Kriegel (2003) for suggestions that accepting local supervenience of phenomenal character, does not necessarily entail that you deny physicalism; only that this local generation of phenomenal character and the resulting experiences must also be eventually described as physical entities. While I accept that this may be harder to explain and conceive of at present, the point here is that this move does not necessarily entail abandoning physicalism as yet.

I appreciate that this remaining 'Generation Mystery', as I refer to it, may seem unsatisfying to some readers[136]. In reply to this, I would remind those readers that this is not a thesis specifically designed to analyse the metaphysics of consciousness. It does involve many elements to do with consciousness, but is mainly focused on looking at the overlap between vision and visual imagery at the sub-personal and personal levels. And since I have already argued for a highly Conjunctivist Position, I may now perhaps feel justified in assuming that the generation process in both vision and visual imagery will be similar in both cases. This is because I think they are underpinned by similar processes and lead to *similar kinds* of locally supervenient phenomenal experiences. The differences I predict are therefore mainly in the ways these mental states are initially caused and eventually experienced (see later), with much of the in-between processes being used in common.

Having said that, there are various theories about the neural correlates of consciousness (NCC) currently on the market, all of which I think are generally compatible with my theory. That is, I am not aware of any theory in this field that needs to be specifically addressed that would potentially contradict my approach[137]. To illustrate this point, here are two theories that are perhaps at opposite ends of the spectrum of the NCC debate. These either predict that a relatively small part of the brain is sufficient for supporting phenomenal experience, or that it is a much more widely distributed event. In terms of the former, they might suggest that the NCCs are in the visual cortex *itself* and these somehow generate phenomenal character on their own. Examples of these kinds of theories are:

---

[136] It is worth noting that I have criticised Tye (1995, 2000) in the last chapter for denying that an explanation of how the signal to phenomenal gap is crossed is required. This is different from my position because I accept that some explanation is required and it is only that I think it is currently premature to attempt to give one. Hence I readily identify this as one part of my theory that remains unexplained, which is importantly different from denying it needs any explanation.

[137] For a fuller survey of the potential neural correlates of consciousness see Chalmers (1998, 2000) and Crick and Koch (1998).

ffytche's (2003) 'micro-consciousness theory' that indicates that activity in visual cortex in general may be sufficient; and Tootell et al (1995) who have more specifically suggested sub-area V5[138]. Conversely, it may be the NCCs are a more distributed global event based on the rhythmic firings of different parts of the cortex, which in turn can be detected as 'brain waves' pulsing at different frequencies. An example of this is Crick and Koch's (1990) 40-hertz oscillations *over the whole cerebral cortex*, which not only includes the visual cortex, but the frontal, parietal and temporal cortexes also. Despite these contrasting claims I would suggest that these bipolar opposite approaches are both generally compatible with my theory and by inference I think all the less extreme ones in between will be to. This is because I suggest the options described above, and any intermediary ones, could all potentially access information from the visual buffer to delimit what is finally experienced after further (same level or upstream) processing. So in terms of what I want to say next it therefore doesn't really matter too much which of the current theories of the NCCs eventually becomes more established, since my theory seems to potentially compatible with them all at present.[139]

The main thing to stress at this point is that I think that the online generation process, what ever it turns out to be, will presumably also be available to generate offline mental imagery. I think this is assured because there is no reason I am aware of, that the other necessary upstream processes couldn't access the visual buffer the same way in both cases. That is they could access information from the visual buffer irrespective of whether it is labelled as online or offline. And the offline 'mock' label could keep imagistic representations isolated from online perceptual representations in further processing. So I

[138] Recall the visual buffer is thought to be in downstream areas V1 and V2 at the most. So the relatively upstream area V5 could easily access these areas as predicted by my theory.

[139] But again mainly for the record and the interest of the reader, I tend to favour a more in-between approach like the one put forward by Kriegel's (2007b) in his paper on this subject. Here he broadly suggests that parts of the frontal cortex access more downstream areas, which also fits with the Kosslyn Model as discussed.

will assume this is at least plausible in what follows, even though admittedly very little is known or has been said about this aspect at present. I will however return to address this point further after the next section.

7.3.2 – Similar kinds of representations leading to different sorts of experiences

So assuming from now on that some form of (admittedly still mysterious) generation processes is involved in somehow generating visual phenomenal experiences in all cases, perhaps this just leaves an explanation of how these overlapping underlying processes give rise to such noticeably different experiences like perception and imagery. In other words perhaps I can now focus on the far right output end of Figure 7.4. What then might underpin the phenomenological difference between apparently experiencing objects externally in perception and experiencing them not-externally during imagery[140]? Well in this respect maybe we can talk about it in terms of where the object is represented as being. In the perception case it is represented as being in the external world and hence you experience it as apparently out there. Perhaps we can think of this as the representation being given specific online egocentric coordinates[141] of where it refers to in real world space. For instance, a red-rectangle on the left five metres away, is represented phenomenally by taking the contents of the visual buffer and conjoining them with its actual calculated external egocentric coordinates. This then controls where the apparent experience is perceived to be occurring and gives rise to an apparent experience as-if of a red rectangle actually out there in the external world five metres away on your left.

---

[140] Recall I introduced and talked about the Perky (1910) experiments to motivate the idea that it is commonly accepted that imagined objects are not experienced externally in the same way that perception is. Hence this is one of the initial assumptions of the thesis and I refer to visual imagery simply as not representing things as being in the actual external world while leaving it open as to how this is developed further.

[141] Recall: that egocentric coordinates locates objects relative to where you are and also locates where you are in the world; and allocentric coordinates only refer to the location of objects relative to other objects, or to the relative location of pertinent features within an object (like the equal distances between the corners of a cube).

In terms of imagery then, maybe we can just give a negative version of the online visual case. This could be done by simply **not** merging the imagistic representation with any actual *online* external egocentric coordinates. That is, in imagery the imagined features are just represented as being viewed from a certain point of view, but this does not require them to actually be representing anything *actually out there now*. Hence there is no need to allocate or merge these with *real online* egocentric coordinates at all, and hence no need to experience them as actually being apparently out there in the external world. In this sense perhaps just getting the relative coordinates of the objects right will do the job for imagery, because this will set the relative locations and sizes of the objects, as delimited by what you remember or imagine that you would have seen from a particular point of view (and see below for a worked example of this). Of course imagining what things look like from a particular position may involve working out your imagined egocentric location relative to them and where you are in the world, but these values need not then be treated as real enough to be experienced as actually happening in the real external world now. So even if in some cases images do need to be allocated some egocentric coordinates, then these can be labelled as merely mock or pretend ones. Hence we just need to imagine the relevant objects and their relative locations and sizes from the imagined viewpoint and this need *not be given any real online external coordinates in any situation*.

So perhaps I can summarise this step in my theory by saying that: whilst in the perceptual case you represent real objects and experience them as being in the external world; in the imagistic case you just represent them as being seen from a particular, possibly standard, imaginary point of view. Or alternatively: whilst in perception it is important to represent the actual egocentric location that the current stimulus is coming from; in imagery you just need to represent things as being seen from a particular imagined (or remembered) viewpoint, and this can be done without representing them as being anywhere in particular in your current environment. Hence in the imagery case *online* egocentric coordinates just aren't necessary for this process, as long as the relative positions of the objects are set properly and the details of the correct viewpoint filled out appropriately, then a non-externally experienced apparently 3D image can be generated and examined. This means that the imagined scene can mimic (or simulate) what you would have seen in the real case, without needing to place it as necessarily being anywhere specific in the external world.

Another related representational difference between imagery and perception is that in imagery there is no need to always reproduce a full and detailed equivalent of the visual scene. This is because it may be more efficient to produce just enough detail in imagery to complete the task at hand. So you can save energy (i.e. brain processing power) by just representing the minimum amount needed, for the minimum time needed, with the minimum vivacity required to complete the job. In other words, why bother generating a fully phenomenal vivid external experience, when sometimes all that is required is a weakly experienced, fleeting, vague point of view on an imagined situation. Although of course this will still have to have just enough detail so you can check the relevant details relevant to solving the problem. This gives an additional efficiency based explanation of the phenomenological differences between a seemingly complete vivid perception and an often relatively less vivid and incomplete image that only persists for as long as it is useful and is focused upon.

It may be helpful at this point to illustrate these ideas with a worked example. Take my by now familiar example of when you imagine a Rubiks Cube and need to count off the small squares from its faces. It seems the visual image of a Rubiks Cube doesn't need to be represented as being anywhere in particular in the environment to do this, you just need to make sure it is experienced with the right shape, size, vivacity and detail, so that you can check how many small squares it has. The phenomenology will be *similar* to the perceptual case, because you are taking a perspective on something that seems shaped and coloured. However in the imagined case this is just not represented as actually occurring in your current external environment and hence not experienced as such. My theory would explain this experiential difference as being due to the fact that the visual image of a Rubiks Cube is not merged with online real egocentric coordinates and hence only experienced from an imaginary, usually default, point of view (e.g. near distance directly ahead). This lack of merging with the online egocentric coordinates means it is never treated as actual, but the details of the image are presented in enough details, in the correct format (apparent shapes and colours) to be able to solve the problem. This therefore gives one way of explaining the differences between perceiving a Rubiks Cube online and imagining one offline. The above characterises my main claims about how different kinds of representational styles might lead to the phenomenal differences we experience between vision and visual imagery in terms of its vividness, completeness and where it is represented as occurring. Below I will show how these claims can be illustrated further and also describe some other features that I think my theory is now compatible with and indeed motivated by in the first place.

To that end perhaps another useful way of thinking about these distinctions is to link this explanation with motor actions. Consider that in online visual perception, it helps enormously if our visual states match up with our tactile *bodily feelings* (as discussed in Chapter 2). This is because we can then use vision to guide actions and to avoid current environmental danger that might damage our bodies: e.g. grabbing things and avoiding lethal drops. So it makes sense to have visual experiences match up with where we experience them bodily, and that means matching them up with the online egocentric coordinate system used in motor control of physical actions, which do to occur in real 3D space. In imagery this is not required and perhaps even needs to be inhibited, because we don't really want to waste energy reacting to what we have merely imagined: e.g. running away from an imagined tiger. So it makes sense to make sure imagery is not experienced as external and is instead just represented as not occurring in the current external environment. Consequently it could have evolved to be specifically isolated from being integrated with the actual online egocentric coordinate system.

I think this approach is compatible and specifically motivated by my explanation of these features in Chapter 4, when I first introduced the egocentric coordinate terminology and talked about the interactions of the OPP and SPP streams in terms of Neglect (visual and motor). It is also highly compatible with Hurley's (2008) Shared Circuits Model that links motor control with visual perception in terms of sensory-motor feedback as discussed in Section 2.3. So there seems to be some potential for interlinking the ways these different sense modalities (e.g. touch and vision) and their respective online and offline versions (e.g. perception and imagery) can work together. And as long as we can keep tight control of what's real and what is imagined, then these varied systems can run quite closely together. For example: by vision and touch identifying a real world problem; and then by using imagery to covertly practice or choose between a range of possible reactions to that problem; and then to use vision and motor control to help guide the action you eventually decide to execute in the external world. This is a cycle that can be repeated indefinitely in order to improve our online decision making process so we can survive and thrive.

It is worth noting here that what I am describing is different but related to using a map-like representational capacity to control what you might *imagine* seeing from a particular viewpoint: as in the Piazza Del Duomo example in Chapter 4. Here you imagine an egocentric geographic location and imagine/remember what you might see if you were standing there and use this to delimit the information that is eventually sent to the visual buffer. However there then seems to be no need for these guiding egocentric geographic coordinates to be now sent along to the visual buffer as well. But even if they are, they can perhaps at least now be labelled as offline and hence only imaginary, so that you don't suddenly think you are actually positioned there.

In the above I have offered one way to think about how the phenomenological differences between perceptual and imagistic states could be explained. Perhaps this then offers a tangible way to account for the differences between externally experienced online vision and visual imagery that is not experienced externally but is still perspectival. And I think it can also account for how imagery is experienced in a vision-like way, but perhaps using only allocentric or imagined-egocentric coordinates, instead of being matched up with actual online external egocentric coordinates. This concludes my discussion of the features illustrated on the right hand side of Figure 7.4. In the next section I will relate this more generally to other theories of phenomenal consciousness before tying this all together in an interim conclusion.

7.3.3 – Visual imagery and Generative Approaches to consciousness

Given the above the reader may now still want to ask: 'If in visual imagery the imagined object is not experienced as being in the external world then where does the image of a Rubiks Cube occur' To which I would reply that in some ways this question has missed the point. The image doesn't occur anywhere as-such, it is an experience that you go through if you happen to be undergoing the appropriate kind of brain processes. That is, it is nowhere but as a physical brain process, involving all the areas discussed above, that gives rise to this apparent 3D phenomenal experience. The question misses the point because it seems the same question can be asked about the apparent externality in the *online perceptual* case, which is also just a locally supervenient mental state, but one that is experienced as apparently external. So an equally pressing question would be to ask how can the brain can generate phenomenal experiences of apparently external 3D objects? But to ask about this feature, is again I would say just to ask about the Generation Mystery.

This is because it seems me that it is also difficult to explain how something like a BIV can locally support an experience that feels like we are just seeing the external world directly as our naïve intuitions suggest. The same difficulty also arises if you accept the standard account of visual hallucination and after-images, which are both locally supervenient vision-like states that give one an experience of an object that is apparently located in the external world, when of course no such object actually exists. So just as you can ask this kind of question about visual imagery, one can ask a similar thing about other vision and other vision-like states. Therefore I would suggest a question about where one experiences a visual image, is just a sub-set of the Generation Mystery question. And I have dealt with this point above by arguing that brain states must somehow be producing these apparently spatial experiences, but we just don't know how as yet and so it remains a mystery. I hope my approach can someday be linked to research that may make some progress in this area, but I think it is too soon to attempt this at present; and this is especially true in the restricted context of this thesis[142].

To put this point another way, I would suggest that once you have accepted that the brain can locally generate phenomenal character in the perceptual case by representing it in apparent shapes and colours in the external world, then it seems to me that this explanation

[142] But I think investigating things like mirrors, 3D cinemas and Magic Eye pictures, can give us some insight how the mind generates 3D experience from what is actually information coming from a flat surface. A good start to this thought process is perhaps to just look at a mirror and think how strange this experience is of you at a spatial point somewhere behind the wall. Now ask yourself: where does that occur? See also Levine (2008) for a related discussion of virtual space.

is also potentially available to explain the phenomenal character of visual imagery. The difference comes not in the generation step, but in where the real or imagined objects are represented as being: where the former are represented as occurring externally and the latter as just represented from a certain (possibly default) point of view. Hence the most difficult step to justify in my opinion, is at the point of accepting local supervenience of phenomenal character, and once this is done we are free to suggest that the brain can represent things in different ways. This is why I spent a significant amount of time developing the Time-Spread Argument to motivate my Generative Approach to phenomenal character.

To now demand details of how this difference in generation occurs is to ask for a detailed account of the way different kinds of visual phenomenal consciousness are generated based on these different kinds of online and offline representations. That is I would need to develop a theory about the generation of consciousness and how it varies among vision and vision-like states. And although this is something I would like to develop in future, I do not have the space to enter into that discussion here in enough detail to do it justice. Instead I suggest it mainly regards a sub-distinction within a general Generation Mystery question and this is something that I have already indicated I will not attempt to do here, because it would be very speculative given our current lack of knowledge of this generation process. Hence I feel that by arguing in detail for a Generative Approach based on the Time-spread Argument, and by providing as detailed an argument from both sides of the signal to phenomenal gap as I think is currently possible, I have done what I think is reasonably possible in order to fulfil the aims of this thesis.

What I we can say that is a bit more positive is that according to the local supervenience thesis, if you are undergoing the appropriate kinds of brain processes, then you will be undergoing a imagistic phenomenal experience somehow, and this doesn't entail that it is going on anywhere apart from as activity in your brain. This may therefore be compatible with the straight forward claims of Identity Theory on the one hand. Or at the other extreme it may be compatible with a more complex theory about the relation between physical matter and consciousness such as Emergentism, where apparently 3D visual experience somehow emerges from the complexity of the neural interactions involved[143]. I can stay neutral in this respect also, as long as this process is equally accessible for taking offline for imagery. I can see no reason why it wouldn't be because it is just a difference in representational style rather than invoking anything ontologically new. And note that the above is not the same as saying prejudicially that this is the only way that this kind of experience can occur. My position is consistent with saying that the same phenomenal experience could be had by many different brain states, or even other physical states with very different supervenience bases (e.g. silicon). The claim here is *only* that if you are in this particular brain state (or better, if *you are that* particular brain state), then you will also be undergoing this kind of phenomenal experience: to deny this under my approach is just to deny materialism.[144]

As mentioned earlier I take my theory to be compatible with any Generative Approach that suggests that the brain generates phenomenal character in all cases. These are generally also referred to as (highest) common factor approaches in the literature (c.f. Crane 2005: Section 3.4) and these usually accept the local supervenience of phenomenal character and try to remain compatible with the BIV intuition. For some examples of theories of consciousness that my theory may be compatible with, I would suggest that Wright's

[143] For Emergentism see for example Crane (2001: Section 18), Chalmers (2006) and Harre (2006). For papers on Identity Theory see for example: Smart (1972) and Borst (1970).

[144] See Kriegel (2002b: 185) for a similar comment: '…it is difficult to conceive of a scenario in which qualia are inverted but the neurophysiological underpinnings of experiences remain the same. In fact, the supervenience of qualia on matter excludes the possibility of such a scenario. To insist that such a scenario is metaphysically possible is to deny qualia-matter supervenience and effectively renounce materialism.'

(2008) compilation called ‘The Case for Qualia’ contains a good representative collection of modern theories that I think are compatible with mine; perhaps also given some relatively minor qualifications. Indeed many of these theories explicitly indicate that they are indirect realist theories and often also appeal to science based arguments in a similar way that I do. However as discussed in Chapter 1, I would prefer to avoid the use of the term ‘qualia’ in the narrow sense[145] as non-intentional states (c.f. Block 1996), since I remain sympathetic to the idea that qualia can represent things and hence can be intentional (c.f. Loar 2003, Levine 2008). An example of this kind of approach, in the Wright (2008) collection, is a paper by Graham and Horgan (2008) on ‘Qualia Realism’, which represents another paper on their continued promotion of a position called Phenomenal Intentionality. Amongst other things this school of thought claims that all qualia are intrinsically intentional states, which is very different from Block’s narrow use of the term and I think this amply illustrates my reluctance to use this term since it is obviously potentially confusing.

Although the broad category of Generative Approaches stated above does include sense-data theories, and I am *relatively* sympathetic to these theories, I tend not to favour these kinds of approaches since they *risk* introducing ontologically separate entities with strange properties into their explanations: i.e. mind independent sense-data objects[146]. I see my approach as most compatible with the recently emerging phenomenal intentionality

[145] And recall in the last chapter I suggested that Wright (2008) only reluctantly uses this term since ‘Representationalism’, his preferred term, has become associated with theories hostile to qualia (e.g. Tye 1995, Dretske 1995)

[146] Hence I would prefer something more in the *spirit* of Adverbialism. But having said that I wouldn’t necessarily take being labelled as a sense-data theory as a very strong criticism taken on its own. This is because I agree with Maund (2005: 40), who suggests that some researchers seem to dismiss sense-data theories far too easily for a ‘…whole host of familiar reasons’ (c.f. Tye 2000: 45-46). He also suggests that the recent work on sense-data theories by Howard Robinson (1994) might deserve a different and more considered treatment. So I wouldn’t necessarily count all sense-data theories as disproved, even though they may be currently out of favour. See also Howell & Fantl (2003) and Hilbert’s (2004) for defences against other criticisms usually levelled against “Highest Common Factor” approaches: e.g. the ‘speckled hen’ and ‘closely matching samples’ objections.

movement[147] because those theories seem to explicitly appeal to the BIV intuition and to endorse a form of non-reductive representationalism. In other words I think it can potentially be classed as both a Representational and Generative Approach, and therefore it seems broadly in lines with my preferred approach to perception and hence imagery[148]. Unfortunately I do not have time to go into the details of this compatibility here, and only mention it here as one suggested way that my theory could be merged and developed along with other more established approaches to perceptual consciousness. It would only be a problem for my approach if all of these kinds of theories were shown to be implausible, but for now there seem to be many live options that I can appeal to in this respect.

7.3.4 – Interim Conclusion on the nature of visual imagery

So in conclusion to this section I would offer the following summary of how I would describe the similarities and differences between perception, hallucination and visual imagery based on the above arguments. [And note in what follows I treat illusion as a variety of perception, all be it a mis-perception, even though I have said little about this so far, and will say little to justify it further; I include it here merely for completeness[149]]:

---

[147] C.f. Horgan and Tienson (2002), Horgan and Kriegel (2008), Horgan et al (2004), Farkas (2008) and Georgalis (2003).

[148] I think my theory aligns particularly closely with the body of work put forward by Kriegel (2002ab, 2003, 2007ab, 2008), especially his work on the NCCs (2007b - as discussed above) and his paper on 'Intentional Inexistence and Phenomenal Intentionality' (2007a), which I see as another way of approaching phenomenal imagination and which defends an Adverbialist approach. See also footnote 152 below for an indication of my theories compatibility with Crane's (2001) approach to intentionality.

[149] But see Thau (2004: 248) and Chapter 5 footnote 80 for more on this.

1) In **visual perceptual states** (including illusions as mis-perceptions) - you are visually representing things about your environment by generating phenomenal experiences of apparently external objects based on the (mostly) externally caused and filled-in *online* contents of the visual buffer.

2) In **visual hallucinations** and **afterimages -** you are visually mis-representing things about your environment by generating phenomenal experiences of apparently external objects based on the (at least partially) internally caused and filled-in *online* contents of the visual buffer.

3) In **visual imagery -** a phenomenal image is generated of a scene from a point of view that simulates at least partly what you might see if this situation were to actually be occurring. It can represent a situation that either: you may have actually already seen (memory); or you could have possibly seen (counterfactual past); or one you might possibly see (future possible). However in this vision-like representation the objects are **not** experienced as being in the external world, but are only experienced from a certain point of view in apparently spatial shapes and colours, and this is based on the internally caused and filled-in *offline* contents of the visual buffer.

AND: in all cases of perception *and* hallucination *and* imagery, they are relatively similar kinds of mental states as they are all *locally generated visual style phenomenal representational mental states.* That is they share a significant common factor in the kinds of phenomenal mental states they are and the crucial brain mechanisms that underpin them; especially in terms of their use of the visual buffer. However there are some minor processing and phenomenological differences as stated above, the main one being that the imagery case is an offline simulation of the online visual case as currently compatible with my interpretation of the Kosslyn Model of perception and imagery.

From the above you will note that I have put this in the form of a three-pronged distinction, which are all brought together as similar kinds of things as described by the last paragraph that starts with ‘AND’. This indicates that this is the most important element and explains the predominant common factor between these states. Specifically this conjunction suggests they are all locally generated visual style phenomenal experiences with significant overlaps in the underlying processing mechanisms they use. This is especially in terms of their crucial use of the visual buffer as suggested by the Kosslyn Model. The idea of the three-way sub-distinction is that it can explain the less crucial phenomenological and processing differences by explaining how perception, hallucination and visual imagery are caused and experienced differently, even if they are ultimately very similar kinds of things.

So the above aims to give a summary explanation of how visual and vision-like states can seem similar in certain ways and have a significant processing overlap, which is in line with our phenomenological experiences and the available empirical data respectively. Hence I hope to have given one way of explaining both the difference in phenomenology between vision and visual imagery, but also the similarities between them based on sub-personal processes and their common factor as locally generated visual style phenomenal experiences. So if this way of putting it is at least plausible, it gives one way of explaining the underlying *nature* of visual imagery and its relation to perception, in line with one of the main aims of the thesis. We can now perhaps apply these points to a final exposition of my full positive account of STIm next.

7.4 – From generic visual imagery to full blown STIm

Although the interim summary above explains certain things that characterise visual imagery in general, it doesn't explain how this comes to be associated with other spatial and temporal contexts. Recall that the above does control for how imagined objects are distributed in apparent 3D space and how their shapes and colours are allocated. This gives them what I have referred to as their *spatial layout*, which I define as the distribution of these imagined objects within an imagined scene. Towards the end of the last chapter (Section 6.2) I developed a labelling-based approach to imagery, which suggested a method by which visual images could become associated with different temporal contexts and/or be imagined to be located in different spatial locations. This involved appending these images with a label as to various aspects that the image was thought to represent. For instance representing an event that may occur in the near future in a different location to the one you are currently occupying. At the end of the last chapter I promised to return to this spatio-temporal labelling aspect of imagery and propose a way that it could be annotated. This is what I intend to do next as a way of concluding my explanation of the full nature and function of STIm in terms of visual imagery.

I think that perhaps the best way of doing this is to provide a worked example to illustrate how an imagined scene with the same spatial layout can come to be thought of under different spatio-temporal contexts. To that end I will return to use my trusted example of a Rubiks Cube and show how a visual image of it can be treated in very different ways depending on how it is labelled. In order to make this as close as possible to a real life situation I have included examples of what your belief and desire attitudes towards this image might be and how these may affect your subsequent actions. These in turn can then be interpreted as illustrating a way that this image could be function in our wider mental economy. Hence I will introduce four cases below where I not only vary the spatio-temporal context of the same Rubiks Cube image, but I also vary your intentional belief and desire attitudes towards it as well.

In order to present these four cases I will annotate the visual image of the Rubiks Cube in exactly the same way in each case. Based on what I have said previously I would annotate the representation of spatial layout of the Rubiks Cube in the visual buffer like this:

Offline (“This represents a Rubiks Cube”) {Array Cell (X, Y): colour (RGB: P, Q, R), depth Z, …etc}

Where the right hand side within the curly brackets refers to the information contained within the visual buffer that represent the spatial parameters of the Rubiks Cube. Since this is a willed episode this has also been appended with the label 'This represents a Rubiks Cube'. This in turn can also be labelled as online or offline depending if these spatial parameters are perceived or imagined respectively. In the example above it is offline and hence represents an imagined Rubiks Cube. This information would then need to be accessed appropriately by the upstream generation processes before a phenomenal image of an apparently shaped and coloured Rubiks Cube could be experienced. However since I want to focus on the labelling (left hand) side of the annotation above, and the right hand side will be the same in each case, I will abbreviate the right hand side to simply this '{VB=RC}', which will indicate that the visual buffer (VB) contains the information needed to visually represent the spatial parameters of a Rubiks Cube (RC) and is also labelled 'This represents a Rubiks Cube'. This will allow me to append more labels to the left hand side in a clearer fashion. Hence in what follows {VB=RC} will represent the information contained in the visual buffer that is accessed and labelled appropriately in order to generate a phenomenal experience as of a Rubiks Cube and will remain constant in each case.

This now allows me to focus on varying the left hand side labelling aspect of this image. So in the list below I will illustrate three ways to annotate the same image and one way to annotate a similar episode of perception. They all represent a different situation that is described in a sentence above each annotation:

Case 1: This represents you imagining a Rubiks Cube that you want to buy:
(Believe= yes, Desire= yes) (Location= toyshop, Time=future/present, possible) –Offline {VB=RC}

Case 2: This represents you imaging a Rubiks Cube you wish you'd had as a child:
(Believe= no, Desire= yes) (Location= non-specific, Time=past, possible) - Offline {VB=RC}

Case 3: This represents you remembering a Rubiks Cube you regret buying on holiday:
(Believe= yes, Desire= no) (Location= Spain, Time=past, actual) – Offline {VB=RC}

Case 4: This represents you actually seeing a Rubiks Cube you want to buy:
(Believe= yes, Desire= yes) (Location= 5m infront, Time=present, actual) - Online {VB=RC}

Table 7.1: **Case 1-3** – are nested annotations of different spatio-temporal contexts and intentional attitudes to the same Rubiks Cube image. **Case 4** – is a nested annotation of a perception of a similar Rubiks Cube. NB – in all cases {VB=RC} is short for the visual buffer being filled and labelled with the relevant representations as of a Rubiks Cube

From the above we can see that in **Case 1-3** our attitude to the same image can vary according to our belief and desires for the object represented in the image. It can also vary in the spatio-temporal context it is labelled as being in. These vary in the temporal domain between: being in the past and factual (memory); to being in the past and counterfactual (imagined); and being a possibility in the future or present (imagined). Hence this covers the four main temporal settings that I think are possible (present, future, past actual, past possible) and shows how the temporal aspect can be varied and controlled. Similarly I have given some examples of how the spatial context of the image can vary. For example: from the local toyshop, to a holiday location in Spain, to being left unspecified. All this even though the actual spatial layout of the image of the Rubiks Cube remains the same in each case. I also see no reason why the spatial location could not be set in non-earth-bound locations, such as Mars, or even in imaginary places (e.g. Narnia).

And recall that in Section 6.2 I have made the assumption that a normal modern adult human subject has the concepts and mental equipment necessary to understand and interpret what these spatio-temporal labels might mean. So although I couldn't go into too much detail about how this aspect actually works, I have suggested this might be a good platform from which an investigation into this aspect of our mental lives might start from. It is also worth noting that varying of spatio-temporal context of the image may in turn change other attitudes you have towards the Rubiks Cube visualization, such as emotional ones of positive happy anticipation or negative sad regret. I have tried to suggest this in the desire aspect above and this is also illustrated in the sailor example I described in Section 6.2.2.

I have included **Case 4** above as a way of contrasting the imagery case with that of perception, although I don't intend to say much more about it here. This is because I think I have discussed how I see the relationship between perceptions and imagery in enough detail already. The main purpose of introducing Case 4 here, is mainly to illustrate how the offline label would be dropped or replaced by an *online* label in perception, and hence how this would mean that it would be represented in consciousness as being actually currently occurring in the external world[150].

So leaving the perceptual case aside for now and returning to discuss the imagery cases further, it seems to me that how the Rubiks Cube image is utilised in your wider economy is really only limited by your imagination as to what the Rubiks Cube can mean to you. For instance: it can retain a fairly factual meaning and represent an annoying puzzle; or become very fictional and creative and come the represent, for example, a key to a magical portal. It all depends on whether you are using it factually to solve problems and guide real

[150] And note I am aware that this may involve changing the information held within {VB=RB}, for instance by filling in more detail, but have left this factor out to simply focus on the labelling process here.

life decisions, or are using it more creatively in an episode of play or fiction. Any more creative imaginative act might include also changing the spatial layout of the Rubiks Cube so that it turns into another shape like a pyramid. And while I have been unable to explore this broader and more creative use of visual imagery here, this is also something that is anticipated by this theory and may be developed in the future[151].

Conversely one of my main purposes here *has* been to show why this capacity may have evolved in the first place, and solving practical problems seems like a good way for imagery to earn its keep by increasing our survival chances. To that end I have now indicated at least two practical problems that an image of a Rubiks Cube could be useful for: the first being in problem solving to work out the number of small squares it has; the second being about guiding the decision-making process of what to do next. As examples of the latter, the desire for a Rubiks Cube in Case 1 could lead to you actually going to the shop to buy a Rubiks Cube, and the negative memory of one in Case 3 might prevent you from impulse buying another one. And the combination of these two applications to practical problem solving is what I see as its main function of STIm. This is what I have alluded to throughout the thesis, variously in terms of: strategy testing, problem solving, decision-making and deliberation (to use Hurley’s 2008 term). So this gives an idea of what I see as the main function of STIm and I have suggested in Chapter 1 how this may integrate with, and contribute to, the currently growing fields studying Episodic Thinking and Mental Time Travel that also investigate thinking of other spatio-temporal situations in order to plan actions.

Therefore in the above I think I have offered a way that different labels could be appended to an image depending on not only the spatio-temporal context but also our other

[151] For more on creative imagination see for example: Beaney (2005), Carruthers (2002), Currie (2003), Harris (2000) and McGinn (2004).

intentional attitudes towards it[152]. This gives an indication of how it can be integrated with further thinking based on the consequences and meanings of these further attitudes. Hence suggesting a way that the phenomenal image could be incorporated and function in our wider mental economies. Hence these latter distinctions conclude my exposition of the features that control how STIm is generated, maintained and labelled and how it is utilised in our thinking processes and why it exists in the first place. It may be worth finally reminding you that this exposition has been mainly to do with STIm in its realisation as phenomenal visual imagery and that this does not preclude an exposition that deals with its more semantic and non-sensory aspects. This limitation was only invoked in order to keep the thesis focused and this represents another way this thesis could be developed further in future.

7.5 – Conclusion

The summary above, and the interim conclusion in the last sub-section, represents my main conclusions as to the nature and function of STIm and how it relates to perception. These can be further summarised to provide an overall conclusion by way of specifically answering the question in the title of the thesis. This question was: 'What is the nature and function of Spatio-temporal Imagination? Can it plausibly be explained as an offline simulation of the visual process?' Hence a succinct concluding reply to that would now go as follows:

[152] And note that this way of talking seems to match up with certain aspects of Crane's (2001) intentional theory of consciousness, where he suggest we can consider the same object (or content) under different modes. For example, consider this quote: 'For I may imagine the little restaurant in Capri, and I may remember it. These states of mind are different, but their contents are the same. To distinguish these states, we need to mention the different ways in which they relate to this content: by memory and by imagination.' Crane (2001: 31). This suggests another way my theory might be integrated with a wider and more established intentional theory of consciousness. I have used some of Crane's (2001, 2003, 2006) distinctions to guide my thesis already and it is interesting to note he has recently published an article that accepts that perception has intentional or representational content, but at the same time rejects the idea that this means it necessarily has propositional content. I see my approach to perception and imagery as largely compatible with his combined claims about phenomenal consciousness, including this most recent one.

The nature of STIm in terms of visual imagery is as a *locally supervenient visual style phenomenal representational mental state,* which is generated by the brain based on the contents of the visual buffer. In this respect it is very similar to perception and hence they share a significant common factor in the kinds of mental states they are. However imagery differs from perception in that it is not caused or experienced externally and operates as an offline or mock version of the equivalent online perceptual processes that gives one an imagined perspective on a visual scene. Hence it seems it can indeed be usefully interpreted as being an offline simulation of a visual process, especially given my interpretation of how this could integrate with the Kosslyn Model based on a hybrid depictive representational style.

Visual images represent the spatial layout of an imagined or remembered scene, and they can also be associated with different spatio-temporal contexts. This is done by labelling them appropriately according to this spatio-temporal context, so that our attitude towards the apparent spatial layout of the image can be modified accordingly. The appending of this spatio-temporal label is what turns generic phenomenal visual imagery into full-blown Spatio-Temporal Imagination (or STIm) and what allows it to aid in our ability to plan our actions for the future and in the present, based on our knowledge of the past and how it could have been different. Thus the main function of STIm is to allow us to strategy test, make better decisions and to be aware of our extended spatio-temporal environment.

Appendix. A - Explanation of the Tower of Hanoi Puzzle

The idea of the puzzle is to get from the initial state to the goal state (see Figure.A1) by moving one ring at a time. The ring that is moved must be placed on a peg before the next ring is moved. Only the top ring on any peg can be moved. Only one ring can be moved at a time.

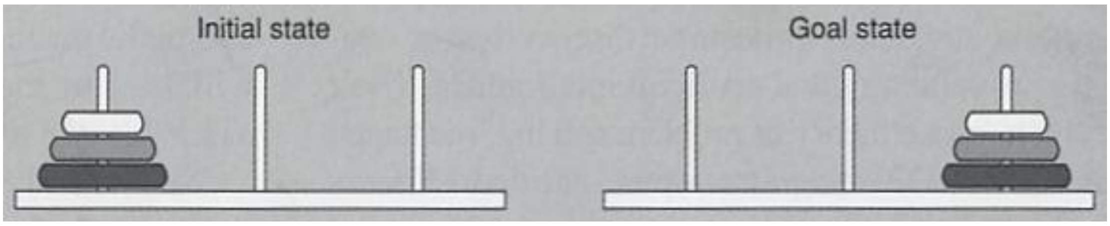


Figure. A1 – The Tower of Hanoi Puzzle – Initial and Goal State

There are a number of ways to the goal state and these are represented below in Fig.A2.

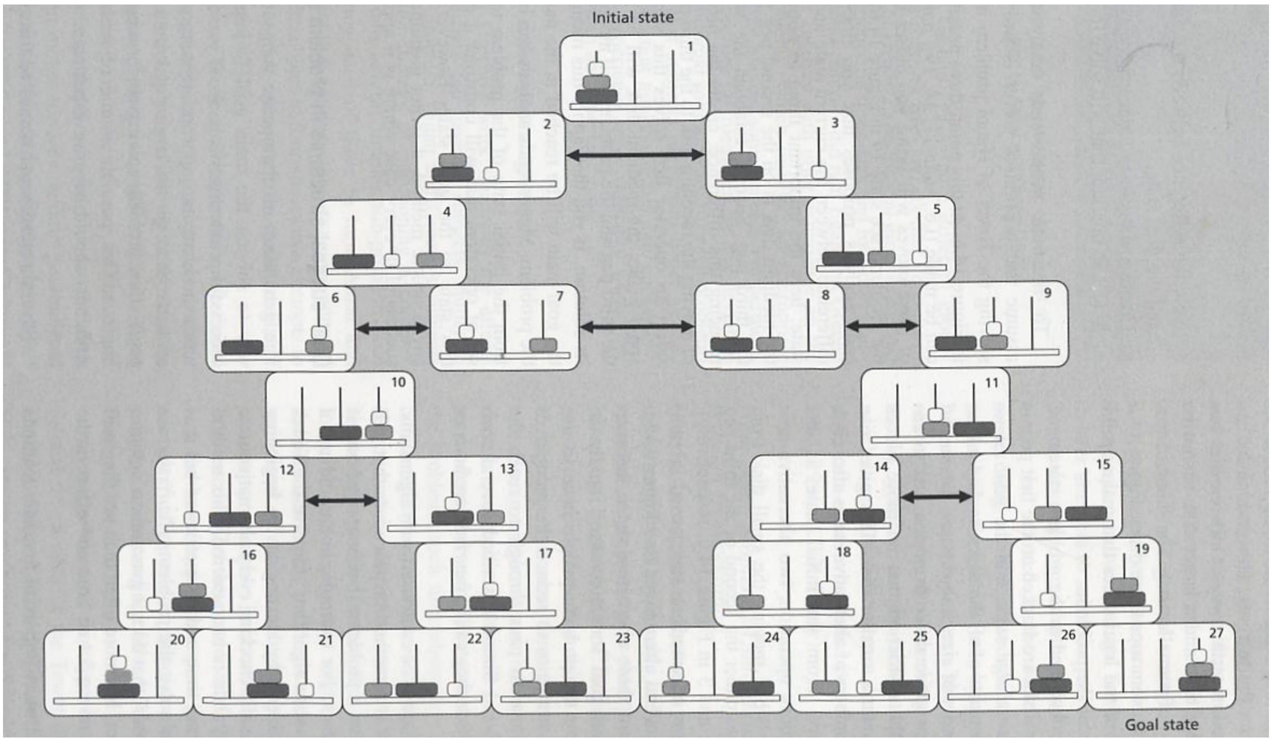


Figure. A2 – Possible routes to the goal state

The Tower of Hanoi Puzzle tests your ability to plan ahead and to remember your past moves. It may also require imagination to maintain an idea of the goal state and compare it with your current progress.

Diagrams taken from: Eysenck, M.W. & Keane, M.T. (2000) – 'Cognitive Psychology: A Students Handbook' *Fourth Edition – Psychology Press Ltd* pgs.400-1

Appendix. B – Main Areas of the Brain

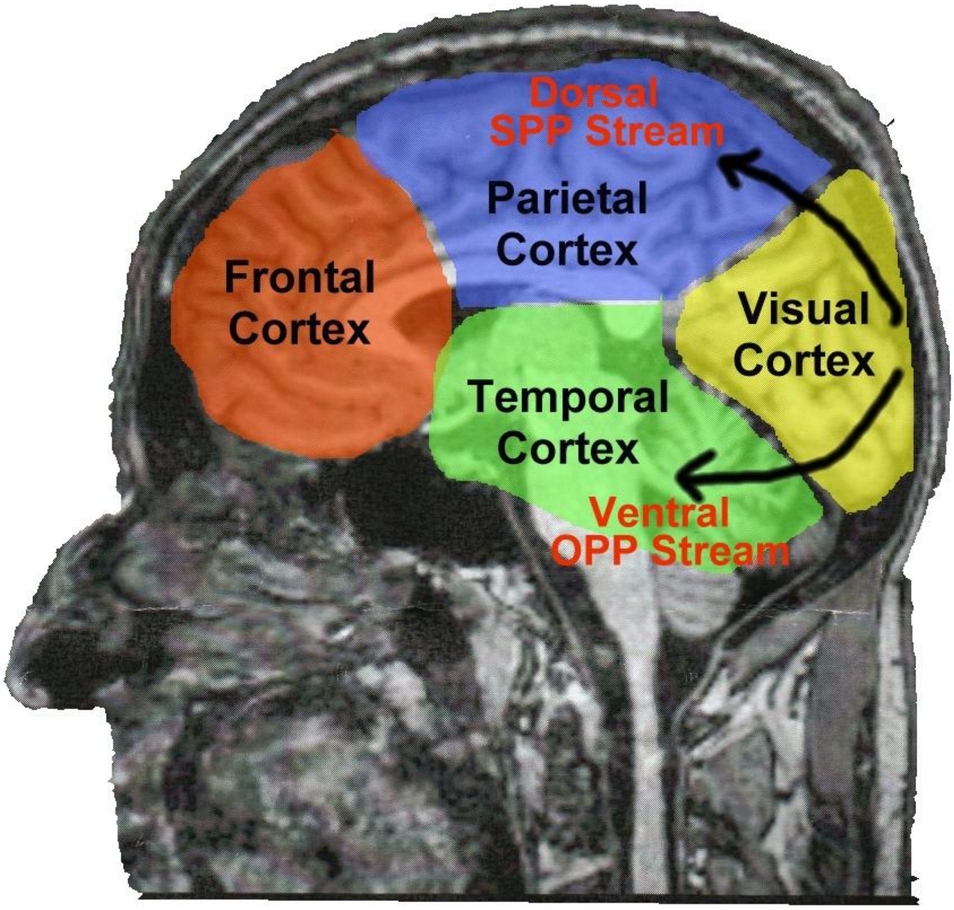


Figure B1 – Approximate locations of areas of cortex and processing streams mentioned in the text – superimposed on a recent scan of the authors brain

References

Alston WP. (1999). Back to the Theory of Appearing. *Noûs,* 33:13 pp.181-203

Alston, WP. (2005). Perception and representation. *Philosophy and Phenomenological Research* 70:2 pp.253-289

Anderson (1978). Arguments Concerning Representations for Mental Imagery. *Psychological Review* 85.4 pp.249-77

Arp, R. (2005). Scenario Visualization: One Explanation of Creative Problem Solving *Journal of Consciousness Studies* 12:3 pp. 31-60

Arp, R. (2008). *Scenario Visualization: An Evolutionary Account of Creative Problem Solving*. MIT

Attance, CM. & O'Neill, DK. (2005). The emergence of episodic future thinking in humans. *Learning and Motivation* 36:2 pp.126-144

Aydede, M. (2004). The language of thought hypothesis. *Stanford Encyclopedia of Philosophy*
Online at: http://plato.stanford.edu/entries/language-thought/ [Accessed Sept 2009]

Barsalou, LW. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences* 22 pp.577–660

Bartolomeo P, D'Erme P. & Gainotti G. (1994). The relationship between visuospatial and representational neglect. *Neurology* 44 pp.1710–1714.

Bartolomeo, P. (2002). The Relationship Between Visual perception and Visual Mental Imagery: A Reappraisal of the Neuropsychological Evidence. *Cortex* 38 pp.357-378

Bartolomeo, P. (2007). Visual Neglect. *Current Opinion in Neurology* 20 pp.381-386

Baylor, GW. (1972). *A Treatise on the Mind's Eye: An Empirical Investigation of Visual Mental Imagery*. Ph.D. thesis, Carnegie-Mellon University, Pittsburgh, PA. (University Microfilms 72-12, 699.)

Beaney, M. (2005). *Imagination and Creativity*. The Open University

Becchio, C. & Bertone, C. (2005). The ontology of neglect. *Consciousness and Cognition* 14 pp. 483–494

Bermudez, JL. (2000). Personal and subpersonal: A difference without a distinction. *Philosophical Explorations* 3:1 pp.63-82

Bermudez, JL. (2003). *Thinking Without Words.* Oxford University Press

Beschin N, Basso A. & Della Sala, S. (2000) Perceiving left and imagining right: dissociation in neglect. *Cortex* 36 pp.401–414

Bisiach E. & Luzzatti C. (1978). Unilateral neglect of representational space. *Cortex* 14 pp.129–133.

Bisiach E, Luzzatti C. & Perani D. (1979). Unilateral neglect, representational schema and consciousness. *Brain* 102 pp.609–618.

Block, N. (1978). Troubles with functionalism. *Minnesota Studies in the Philosophy of Science* 9 pp.261-325

Block, N. (1980). Functionalism: In Ned Block (ed.), *Readings in the Philosophy of Psychology*. Harvard University Press

Block, N. (1981, ed). *Imagery*. Cambridge, MA: MIT Press.

Block, N. (1983). Mental Pictures and Cognitive Science. *Philosophical Review* 92 pp.499-539.

Block, N (1996). Mental paint and mental latex. *Philosophical issues* 7 pp.19-48

Block, N. (1998). Is experiencing just representing? *Philosophy and Phenomenological Research* 58:3 pp.663-670

Block, N. (2003). Mental paint. :In M. Hahn & B. Ramberg (eds.) *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. MIT Press

Bonjour, L. (2004). In Search of Direct Realism. *Philosophy and Phenomenological Research* 69:2 pp.349 - 367

Borst, CV. (1970, ed.). *The Mind/Brain Identity Theory.* Macmillan

Boucher, J. (1996). What could possibly explain autism? :in P.Carruthers and P.K. Smith (eds) *Theories of Theories of Mind*. CUP

Brewer, B. (1992). Unilateral neglect and the objectivity of spatial representation. *Mind and Language* 7:3 pp.222-39

Brewer, B. (1999). *Perception and Reason.* Oxford University Press

Brueckner, T. (2004). Brains in a Vat. *Stanford Encyclopedia of Philosophy.*
Online at: http://plato.stanford.edu/entries/brain-vat/ [Accessed Sept 2009]

Byrne, A. & Hilbert DR. (2003). Color realism and color science. *Behavioral and Brain Sciences* 26 pp:3-21

Byrne, A. & Logue, H. (2008). Either/Or :in A. Haddock and F. Macpherson (eds). *Disjunctivism: Perception, Action, Knowledge*. OUP

Byrne, R. & Whiten, A. (1988, eds). *Machiavellian Intelligence: Social Intelligence and the Evolution of Intellect in Monkeys, Apes, and Humans*. OUP

Byrne, R. & Whiten, A. (1997, eds). *Machiavellian Intelligence II : Extensions and Evaluations*. Cambridge

Buckner, RL. & Carroll, DC. (2007). Self-projection and the brain. *Trends Cognitive Science*. 11:2 pp.49-57.

Carlson, N. (1994). *Physiology of Behaviour*. 5th Edition - Allyn and Bacon

Carruthers, P. & Smith, PK. (1996, eds). *Theories of Theories of Mind.* Cambridge University Press

Carruthers, P. (1996). Autism as mind-blindness :in P.Carruthers and P.K. Smith (eds). *Theories of Theories of Mind.* Cambridge

Carruthers, P. & Chamberlain, A. (2000, eds). *Evolution and the Human Mind*. Cambridge

Carruthers, P. (2002). Human creativity. *British Journal For The Philosophy Of Science* 53:2 pp.225-249

Chalmers, DJ. (1998). On the search for the neural correlate of consciousness :In SR. Hameroff, AW. Kaszniak & AC. Scott (eds). *Toward a Science of Consciousness II*. MIT Press

Chalmers, David J. (2000). What is a neural correlate of consciousness? In Thomas Metzinger (ed.), *Neural Correlates of Consciousness*. MIT Press

Chalmers, DJ. (2003). Consciousness and its place in nature. In SP. Stich & TA. Warfield (eds), *Blackwell Guide to the Philosophy of Mind*. Blackwell

Chalmers, DJ. (2004). The representational character of experience :In B. Leiter (ed), *The Future for Philosophy*. Oxford University Press

Chalmers, DJ. (2006). Strong and weak emergence :In P. Davies & P. Clayton (eds), *The Re-Emergence of Emergence*. Oxford University Press

Churchland, PM. (2005). Functionalism at forty: A critical retrospective. *Journal of Philosophy* 102:1 pp.33-50

Clark, A. & Grush, R. (1999). Towards a Cognitive Robotics. *Adaptive Behavior* 7 pp.5-16

Clayton, NS, Bussey, TJ. & Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nature Reviews Neuroscience* 4 pp.685-691

Coslett, HB. (1997). Neglect in vision and visual imagery: a double dissociation. *Brain* 120 pp.1163–1171.

Crane, T. (2001). *Elements of Mind: An Introduction to the Philosophy of Mind.* Oxford University Press

Crane, T. (2003). The intentional structure of consciousness. In Q. Smith & A. Jokic (eds), *Consciousness: New Philosophical Perspectives*. Oxford University Press

Crane, T. (2005). Problems of Perception. *Stanford Encyclopedia of Philosophy*. Online at: http://plato.stanford.edu/entries/perception-problem [Accessed Sept 2009]

Crane, T. (2006). Is There a Perceptual Relation? :in TS. Gendler and J. Hawthorne (eds), *Perceptual Experience*. OUP

Crick, F. and Koch, C. (1990). Towards a Neurobiological Theory of Consciousness. *Seminars in the Neurosciences* 2 pp.262-275 : also reprinted in Block, Flanagan & Guzeldere (1997, eds), *The nature of consciousness.* MIT

Crick, F. & Koch, C. (1998). Consciousness and neuroscience. *Cereb. Cortex* 8 pp.97–107

Currie, G, (1995a). Visual Imagery as the Simulation of Vision. *Mind and Language* 10 pp.25-44

Currie, G. (1995b). Imagination and Simulation: Aesthetics Meets Cognitive Science :in M.Davies & T.Stone (eds), *Mental Simulation*. pp.151-169 Oxford, Basil Blackwell

Currie, G. (1996). Simulation-theory, theory-theory, and evidence from autism : in P. Carruthers and PK. Smith (eds), *Theories of Theories of Mind.* Cambridge

Currie, G. & Ravenscroft, I. (1997). Mental Simulation and Motor Imagery. *Philosophy of Science* 64 pp.161-180

Currie, G. & Ravenscroft, I. (2002). *Recreative Minds*. OUP

Currie, G. (2003). The capacities that enable us to produce and consume art. In TS Gendler, M Kieran, D Lopes (eds*), Imagination, philosophy and the arts*. Routledge

Davies, M. & Stone, T. (2000, eds). *Mental Simulation*. Oxford

Dennett, DC. (1968). *Content and Consciousness.* Routledge

Dennett, DC. (1978). *Brainstorms: Philosophical Essays in Mind and Psychology*. Cambridge Mass.: Bradford Books/MIT Press

Dennett, DC. (1995). *Darwin's Dangerous Idea*. Penguin

DeRose, L. (1999). Contextualism: An Explanation and Defence :in J. Greco & E.Sosa (eds), *The Blackwell Guide to Epistemology*. Blackwell

Djukic, G. & Popescu, VB. (2003). A critique of Langsam's The Theory of Appearing Defended *Philosophical Studies* 112:1 pp.69-91

Dretske, F. (1995). *Naturalizing the Mind.* MIT Press

Duhamel, JR, Colby, C, & Goldberg, ME. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science,* 255: 5040 pp.90–92

Dunbar, R. (2000). On the origins of the human mind :in P. Carruthers & A. Chamberlain (eds), *Evolution and the Human Mind* Cambridge

Dunn, J. (2008). The obscure act of perception. *Philosophical Studies* 139:3 pp.367-393

Farkas, K. (2008). Phenomenal intentionality without compromise. *The Monist* 91:2 pp.273-93

Faw, B. (2003). Pre-frontal executive committee for perception, working memory, attention, long-term memory, motor control, and thinking: a tutorial review. *Consciousness and Cognition* 12 pp.83-139

ffytche, D. & Pins, D. (2003). The Neural Correlates of Conscious Vision. *Cerebral Cortex* 13:5 pp.461-474

Fish, WC. (2004). The direct/indirect distinction in contemporary philosophy of perception. Essays in Philosophy 5 (1):1-13:
(online) http://www.humboldt.edu/~essays/billfish.html [Accessed Sept 09]

Fish, WC. (2005). Disjunctivism and non-disjunctivism: Making sense of the debate. Proceedings of the Aristotelian Society 105:1 pp.119-127

Fish, WC. (2008). Disjunctivism, Indistinguishability, and the Nature of Hallucination :in A. Haddock and F. Macpherson (eds), *Disjunctivism: Perception, Action, Knowledge*. OUP

Fish , WC. (2009). *Perception, Hallucination, and Illusion.* OUP: New York

Frith, CD, Friston, K, Liddle, PF. & Frackowiak, RSJ. (1991). Willed action and the prefrontal cortex in man: A study with PET. *Proc. R. Soc. London B Biol. Sci.* 244 pp.241–246

Frith, CD. (1992) *The cognitive neuropsychology of schizophrenia*. Erlbaum/Taylor & Francis.
Fodor, JA. (1968). The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy* 65 pp.627-40

Fodor, JA. (1975). *The Language of Thought*. New York: Thomas Crowell. (Paperback edition: Harvard University Press, 1980)

Fodor, JA. (2003). Review of Bermudez's *Thinking Without Words.* The Guardian - Online at:
http://www.nyu.edu/gsas/dept/philo/courses/representation/papers/FodorOnBermudez.pdf [Accessed Sept 2009]

Ganis, G, Thompson WL. & Kosslyn, SK. (2004). Brain areas underlying visual mental imagery and visual perception: an fMRI study *Cognitive Brain Research* 20:2 pp. 226-241

Gendler, TS. and Hawthorne, J. (2006). *Perceptual Experience*. OUP

Georgalis, N. (2003). The fiction of phenomenal intentionality. *Consciousness and Emotion* 4:2 pp.243-256

Goldenberg G, Mullbacher, W. & Nowak, A. (1995). Imagery without perception – A case study of anosognosia for cortical blindness. *Neuropsychologia*, 33 pp.1373-1382

Goldman, AI. (2006). *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading.* Oxford University Press

Goldman, AI. (2008). Hurley on Simulation. *Philosophy and Phenomenological Research* Vol. LXXVII No. 3 pp.775-788

Goodale, MA. & Murphy, KJ. (2000). Space in the brain: neural substrates for allocentric and egocentric frames of reference :In T. Metzinger (ed) *Neural correlates of consciousness*. MIT

Graham, G. & Horgan, T. (2008). Qualia Realism: Its Phenomenal Contents and Discontents :in E.Wright (ed). *The Case for Qualia*. MIT Cambridge

Grush, R. (1995). *Emulation and Cognition.* Dissertation, University of California, San Diego

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences* 27 pp.377-442

Grush, R. (2007). Skill theory v2.0: dispositions, emulation, and spatial perception. *Synthese* 159:3 pp.389-416

Guariglia, C, Padovani, A, Pantano P. & Pizzamiglio L. (1993). Unilateral neglect restricted to visual imagery. *Nature* 364 pp.235–237.

Hacking, I. (1991). The Tradition of Natural Kinds. *Philosophical Studies* 61 pp.109-126

Haddock, A. and Macpherson, F. (2008). *Disjunctivism: Perception, Action, Knowledge*. OUP

Hardin, CL. (2008). Color Qualities and the Physical World. In E.Wright (ed). *The Case for Qualia*. MIT Cambridge

Harré, R. (2006). Resolving the emergence-reduction debate. *Synthese* 151:3 pp.499-509

Harris, P. (2000). *The Work of the Imagination*. Blackwell

Hassabis D, Kumaran D. & Maguire EA (2007). Using imagination to understand the neural basis of episodic memory. J Neurosci. 27:52 pp.14365-74

Heal, J. (1996). Simulation and cognitive penetrability. *Mind and Language* 11 pp. 44-67

Heal, J. (1998). Co-cognition and offline simulation. *Mind and Language* 13 pp. 477-98

Heal, J. (2003). *Mind, Reason and Imagination*. CUP

Heal, J. (2005). Review of Stich and Nichols (2003) 'Mindreading' *Mind* 114 pp.181-84

Hilbert, DR. (2004). Hallucination, sense-data and direct realism. *Philosophical Studies* 120 (1-3) pp.185-191

Horgan, TE. & Tienson, JL. (2002). The intentionality of phenomenology and the phenomenology of intentionality :in David J. Chalmers (ed), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press

Horgan, TE, Tienson, JL. & Graham, G. (2004). Phenomenal intentionality and the brain in a vat. In Richard Schantz (ed), *The Externalist Challenge*. Walter de Gruyter.

Horgan, TE. & Kriegel, U. (2008). Phenomenal intentionality meets the extended mind. The Monist 91 pp.347-373

Hornsby, J. (2000). Personal and sub-personal: A defence of Dennett's early distinction. Philosophical Explorations 3:1 pp.6-24

Howell, RJ. & Fantl, J. (2003). Sensations, swatches, and speckled hens. *Pacific Philosophical Quarterly* 84:4 pp.371-383

Howell, RJ. (2008). Subjective Physicalism :in E.Wright (ed). *The Case for Qualia*. MIT Cambridge

Humphrey, NK. (1976). The Social Function of Intellect :in G. Bateson & RA. Hinde (eds) *Growing Points in Ethology.* Cambridge University Press

Hurley, SL. (1998). *Consciousness in Action.* Harvard University Press

Hurley, SL. (2001). Perception and action: Alternative views. *Synthese* 129:1 pp.3-40

Hurley, S. (2008). The Shared Circuits Model: How Control, Mirroring and Simulation Can Enable Imitation, Deliberation, and Mindreading. *Behavioral and Brain Sciences* 31:1 pp.33-34

Hurley, SL. (2008b). Understanding Simulation. *Philosophy and Phenomenological Research* Vol. LXXVII No. 3 pp.755-774

Jeannerod, M. & Frak, V. (1999). Mental imaging of motor activity in humans. *Current Opinion in Neurobiology* 9:6 pp.735-739

Jeannerod, M. (2001). Neural Simulation of Action: A Unifying Mechanism for Motor Cognition *NeuroImage* 14:1 pp.S103-S109

Johnson, SH. (2000). Thinking ahead: The case for motor imagery in prospective judgements
of prehension. *Cognition* 74 pp.33–70.

Johnston, M. (2004). The obscure object of hallucination. *Philosophical Studies* 120(1-3) pp.113-83

Kamitani Y, Miyawaki Y, Uchida H, Yamashita O, Sato M, Morito Y, Tanabe H & Sadato N, (2008). Visual Image Reconstruction from Human Brain Activity using a Combination of Multiscale Local Image Decoders *Neuron* 60:5 pp.915-929

Kay NK, Naselaris T, Prenger RJ. & Gallant JL. (2007). Identifying natural images from human brain activity *Nature* 452 pp.352-355

Kennedy, M. (2007). Visual Awareness of Properties. *Philosophy and Phenomenological Research* Vol. LXXV No. 2 pp.298-324

Kennedy, M. (2009: forthcoming). Heirs of Nothing: The Implications of Transparency. *Philosophy and Phenomenological Research* – preprint available online: http://matthew.kennedy.philosophy.googlepages.com/papers [Accessed Sept 09]

Kind, A. (2003). What's so transparent about transparency? *Philosophical Studies* 115:3 pp.225-244

King, PR. (2002). What evidence is there that children with autism lack imagination? What, if anything, does this tell us about how important imagination is for human beings? :as part of Philosophy BA 'Imagination' module.

King, PR. (2003). What are the most plausible theories about why imagination evolved in humans? :as part of Philosophy BA 'Imagination' module.

King, PR. (2005a). Is there a plausible theory of how Spatio-Temporal Imagination could have evolved? What part did this play in the early development of Internal Representations?' : as a short essay part of MA Research

King, PR. (2005b). How important is Spatio-Temporal Imagination to the functioning of the modern human mind? : as a short essay as part of MA Research.

King, PR. (2005c). What is the nature and function of Stich and Nichol's (2003) 'Possible World Box'? What role does it play in Pretence and Mindreading and how does this relate to the remit of Recreative Imagination?' :as MA Research thesis

King, PR. (forthcoming) Book Review of: Fish, WC. (2009) *Perception, Hallucination, and Illusion.* OUP: New York – forthcoming in *Philosophical Psychology*

Kosslyn SM, Ball TM. & Reiser BJ. (1978). Visual images preserve metric spatial information: evidence from studies of image scanning. *Journal of Experimental Psychology* 4:1 pp.147-60

Kosslyn, SM. (1978). Measuring the visual angle of the mind's eye. *Cognitive Psychology* 10 pp.356–389.

Kosslyn, SM. (1994). *Image and Brain: The Resolution of the Imagery Debate.* MIT Press, Cambridge

Kosslyn, SM. & Sussman, AL. (1995). Roles of imagery in perception: Or, there is no such thing as immaculate perception. In M.S. Gazzaniga (ed) *The cognitive neurosciences.* pp.1035–42. MIT Press.

Kosslyn, SM, Thompson, WL, Kim, IJ. & Alpert, NM. (1995). Topographical representations of mental images in primary visual cortex. *Nature* 378 pp.496–498.

Kosslyn SM, Thompson, WL. & Alpert, NM. (1997) – 'Neural systems shared by visual imagery and visual perception: A positron emission tomography study.' – *NeuroImage* 6 pp.320-334

Kosslyn, SM, Ganis, G. & Thompson, WL. (2001). Neural Foundations of imagery. *Nature reviews Neuroscience* 2 pp.635-642

Kosslyn SM, Thompson WL. & Ganis, G. (2006). *The Case For Mental Imagery. Oxford Psychology Series*

Kriegel, U. (2002a). PANIC theory and the prospects for a representational theory of phenomenal consciousness. *Philosophical Psychology* 15:1 pp.55-64

Kriegel, U. (2002b). Phenomenal content. *Erkenntnis* 57:2 pp.175-198

Kriegel, U. (2003). Is intentionality dependent upon consciousness? *Philosophical Studies* 116:3 pp.271-307

Kriegel, U. (2007a). Intentional Inexistence and Phenomenal Intentionality. *Philosophical Perspectives* 21:1 pp.307-340

Kriegel, U. (2007b). A cross-order integration hypothesis for the neural correlate of consciousness. *Consciousness and Cognition* 16:4 pp.897-912

Kriegel, U. (2008). Real narrow content. *Mind and Language* 23:3 pp.304–328

Langsam, H. (1997). The theory of appearing defended. *Philosophical Studies* 87:1 pp.33-59

Le Morvan, P. (2004). Arguments Against Direct Realism and How to Counter Them. *American Philosophy Quarterly* 41:3 pp.221-234

Levine, J. (2008). Secondary qualities: where consciousness and intentionality meet. *The Monist* 91:2

Loar, B. (2003). Transparent experience and the availability of qualia :in Q. Smith & A. Jokic (eds), *Consciousness: New Philosophical Perspectives*. Oxford University Press

Lycan, WG. (1981). Form, function and feel. *Journal of Philosophy* 78 pp.24-50

Lycan, WG. (1988). *Judgment and Justification.* Cambridge: CUP

Lycan, WG. (2001). The case for phenomenal externalism. *Philosophical Perspectives* 15 pp.17-35

Lycan, WG. (2006). Representational Theories of Consciousness :in *Stanford Encyclopedia of Philosophy* – Online at: http://plato.stanford.edu/entries/consciousness-representational [Accessed Sept 2009]

Macpherson, F. (2003). Novel colours and the content of experience. *Pacific Philosophical Quarterly* 84:1 pp.43-66

Markosian, N. (2002). *A defence of Presentism*. OUP

Marr, D. (1982). *Vision.* Freeman

Martin, MGF. (1992). Perception, concepts, and memory. *Philosophical Review* 101:4 pp.745-63

Martin, MGF. (1997). The reality of appearances. In M. Sainsbury (ed), *Thought and Ontology*. Franco Angeli

Martin, MGF. (2001). Out of the Past: Episodic Recall as Retained Acquaintance :in C. Hoerl & T. McCormack (eds) *Time and Memory.* OUP

Martin, MGF. (2002). The Transparency of Experience. *Mind and Language* 17:3 pp.376-425

Martin, MGF. (2004). The limits of self-awareness. *Philosophical Studies* 120(1-3) pp.37-89

Martin, MGF. (2006). On being alienated :in TS. Gendler & J. Hawthorne (eds), *Perceptual Experience*. Oxford University Press

Maund, B. (2005). Michael Tye on pain and representational content. In M. Aydede (ed), *Pain: New Essays on Its Nature and the Methodology of Its Study*. Cambridge MA: Bradford Book/MIT Press

McGinn, C. (2004). *Mindsight: Image, Dream, Meaning*. Harvard

McLaughlin, B. (2003). Colour, Consciousness and Colour Consciousness' - in B. McLaughlin, Q Smith & A. Jokic (eds) *Consciousness: New Philosophical Perspectives*. OUP

McLaughlin, B. & Bennett, K. (2005) Supervenience :in *Stanford Encyclopedia of Philosophy* – Online at: http://plato.stanford.edu/entries/supervenience/ [Accessed Sept 2009]

Mellet E, Tzourio N, Crivello F, Joliot M, Denis M. & Mazoyer B. (1996). Functional Anatomy of Spatial Mental Imagery Generated from Verbal Instructions. *Journal of Neuroscience* 16:20 pp. 6504-6512

Mellet E, Tzourio N, Denis, M. & Mazoyer, B. (1998). Cortical anatomy of mental imagery of concrete nouns based on their dictionary definition. *NeuroReport* 9:5 pp 803-808

Mellet E, Tzourio-Mazoyer N, Crivello F, Mazoyer B. & Mazard A. (2004) A PET meta-analysis of object and spatial mental imagery. *European Journal of Cognitive Psychology* 16:5 pp. 673 – 695

Mesulam, MM. (1999). Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philosophical transactions-Royal Society of London. Biological sciences* 354:1387 pp.1325-1346

Millikan, R. (2006). Styles of rationality. In: S. Hurley & M. Nudds (eds) *Rational animals?* pp. 117–26. Oxford University Press.

Milner, AD. & Goodale, MA. 1995. *The Visual Brain in Action*. OUP:Oxford

Moran, TP. (1973). *The Symbolic Imagery Hypothesis: A Production System Model*. Ph.D. thesis. Carnegie-Mellon University, Pittsburgh, PA. (University Microfilms 74-14,657.)

Nagel, T. (1974). What is it like to be a bat?. *The Philosophical Review* 83:4 pp.435-450

New Scientist (2007: 24 March): 'Future Recall' by Jessica Marshall pp.36-40

New Scientist (2008: 20 December): 'How to read images straight from the brain' by Celeste Biever pp.13

Noe A, Thompson, E. & Pessoa L. (1998). Finding out about filling-in: A guide to perceptual completion for visual science and the philosophy of perception. *Behavioral and Brain Sciences* 21 pp.723–802

Noe, A. (2006). *Action in perception*. MIT

Ortigue S, Viaud-Delmon I, Annoni JM et al (2001). Pure representational neglect after right thalamic lesion. *Ann Neurol* 50 pp.401–404

Paivio, A. (2007). *Mind and Its Evolution: A Dual Coding Theoretical Approach*. Routledge

Pautz, A. (2006). Sensory awareness is not a wide physical relation: An empirical argument against externalist intentionalism. *Noûs* 40:2 pp.205-240

Pautz, A. (2007). Intentionalism and perceptual presence. *Philosophical Perspectives* 21:1 pp.495-541

Perky, CW. (1910) An Experimental Study of Imagination. *American Journal of Psychology* 21 pp. 422-52

Pennington, B. & Ozonoff, S. (1996). Executive Functions and Developmental Psychology. *Journal of Child Psychology and Psychiatry and Allied Disciplines* 37 pp.51-87

Perner, J. & Lang, B. (2000) Theory of mind and executive function: is there a developmental relationship? :in S.Baron-Cohen, H.Tager-Flusberg, and D.Cohen (eds) *Understanding Other Minds*. Oxford

Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. MIT

Prinz, J. (2006). Putting the Brakes on Enactive Perception. *Psyche* 12:1

Putnam, H. (1981). Mind and body :in *Reason, Truth and History*. Cambridge University Press

Pylyshyn, ZW. (2002). Mental Imagery: In search of a theory. *Behavioral and Brain Sciences* 25 pp.157-182

Pylyshyn, ZW. (2003). *Seeing and Visualizing : It's Not What You Think.* Cambridge, MA: MIT Press

Rao, RPN. (1999). An optimal estimation approach to visual perception and learning. *VisionResearch 39* pp.1963–1989

Rey, G. (1998). A narrow representationalist account of qualitative experience. *Philosophical Perspectives* 12 pp.435-58

Rizzolatti, G. (2005). The mirror neuron system and imitation. In: S. Hurley & N. Chater (eds) *Perspectives on imitation: From neuroscience to social science*. vol. 1 pp. 55–76. MIT Press

Robinson, H. (1994). Perception. *Routledge*

Rode, G, Revol, P, Rossetti, Y, Boisson, D. & Bartolomeo, P. (2007). Looking while imagining: The influence of visual input on representational neglect**.** *Neurology* 68 pp.432–437

Rosenthal, DM. (2000). Metacognition and higher-order thoughts. *Consciousness and Cognition* 9:2 pp.231-242

Russell, J. (1997). *Autism as an Executive Disorder*. Oxford: OUP

Schacter DL, Addis DR. & Buckner RL. (2007). Remembering the past to imagine the future: the prospective brain. *Nat Rev Neurosci*. 8:9 pp.657-61

Segal, SJ. (1971). Processing of the Stimulus in Imagery and Perception :in S.J. Segal (ed) *Imagery: Current Cognitive Approaches*, pp. 73-100. New York: Academic Press

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, et al (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268: 5212 pp.889-893

Shallice, T. (1997). Modularity and Consciousness :in Block, Flanagan & Guzeldere (ed) *The nature of consciousness.* MIT

Shepard, RN. & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *Science* 171 pp.701-703Sider, T. (2001). *Four-Dimensionalism: An Ontology of Persistence and Time*. Clarendon Press

Shimojo S, Kamitani, Y. & Nishida, S. (2001). Afterimage of Perceptually Filled-in Surface. *Science* 293 pp.1677-1680

Siewert, C. (2004). Is Experience Transparent? *Philosophical Studies* 117:1-2 pp.15-41

Slotnick SD, Thompson, WL. & Kosslyn, SM. (2005). Visual Mental Imagery Induces Retinotopically Organized Activation of Early Visual Areas. *Cerebral Cortex* 15:10 pp.1570-1583

Smart, JJC. (1972). Further thoughts on the identity theory. *The Monist* 56 pp.177-92

Snowdon, P. (2005). The Formulation of Disjunctivism: A Response to Fish. *Proceedings of the Aristotelian Society* 105 pp.129-41

Sosa, E. (2000). Contextualism and Scepticism. *Philosophical Issues* 11 pp.1-18

Stich, SP. & Nichols, S. (2002). Folk psychology : in SP. Stich & TA. Warfield (eds.), *Blackwell Guide to Philosophy of Mind*. Blackwell

Stich, SP. & Nichols, S. (2003). *Mindreading. An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds.* Oxford University Press

Strawson, G. (1994). *Mental Reality*. MIT Press

Sturgeon, S. (1998). Visual experience. *Proceedings of the Aristotelian Society* 72:2 pp.179-200

Suddendorf , T. & Corballis, MC. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs*

Suddendorf, T. & Whiten, A. (2007). Mental evolution and development: evidence for secondary representation in children, great ages, and other animals. *Psychol Bull*. 127:5 pp.629-50

Suddendorf , T. & Corballis, MC. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences* 30 pp.299-313

Thau, M. (2004). What is Disjunctivism. *Philosophical Studies* 120 193-253

Thomas, NJT. (2008). Mental Imagery. : in *Stanford Encyclopedia of Philosophy* Online at: http://plato.stanford.edu/entries/mental-imagery/ [Accessed Sept 2009]

Tootell RB, Silverman MS, Switkes E. & De Valois, RL. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, 218:4575 pp.902-904

Tootell, R.B., Reppas, J.B., Dale, A.M., Look, R.B., Sereno, M.I., Malach, R., Brady, J. & Rosen, B.R. (1995). Visual motion aftereffect in human cortical area MT revealed by functional magnetic resonance imaging. *Nature* 375:139-41

Tulving, E. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology* 53 pp:1-25

Tye, M. (1991). *The Imagery Debate*. Cambridge, MA: MIT Press

Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind.* MIT Press

Tye, M. (1996). Orgasms again. *Philosophical Issues* 7 pp.51-54

Tye, M. (2000). *Consciousness, Color, and Content*. MIT

Tye, M. (2003). Blurry images, double vision, and other oddities: New problems for representationalism? In Q. Smith & A. Jokic (eds), *Consciousness: New Philosophical Perspectives*. Oxford University Press

Tye, M. (2005). In defense of representationalism: Reply to commentaries :in M. Aydede (ed), *Pain: New Essays on Its Nature and the Methodology of Its Study*. Cambridge MA: Bradford Book/MIT Press

Tye, M. (2009). *Consciousness Revisited: Materialism without Phenomenal Concepts.* MIT

Ungerleider, LG. & Mishkin, M. (1982). Two Cortical Visual Systems :in DJ. Ingle, MA, Goodale & RJW, Mansfield. *Analysis of visual behavior*. pp.549-586 MIT: Cambridge

Wiggins, D. (1980) *Sameness and Substance*. Cambridge, MA: Harvard University Press.

Wright, EL. (2008, ed). *The Case for Qualia*. Bradford Books

Wright, EL. (1993, ed). *New Representationalisms: Essays in the Philosophy of Perception.* Ashgate

Wright, W. (2003). Projectivist representationalism and color. *Philosophical Psychology* 16:4 pp.515-529