

**A Critical Approach to Representationalism from a Largely Sellarsian
Perspective**

Vladislav Vojtovic

PhD

University of York

Department of Philosophy

September 2010

Abstract

I examine representational theories of phenomenal consciousness. I first argue that none of the common objections to representationalism threaten qualia's reducibility to intentional contents. Instead, far more insight into this reductive strategy is to be gained if we attempt to capture, in representational terms, the difference between qualitative and non-qualitative perceptions. The challenge, when presented for first-order representationalism, calls for a formulation of a higher-order condition. However, if Sellars is right about the nature of our introspective conceptions regarding experiences, the relevant higher-order thoughts do not contain concepts which distinctively apply to qualitative, as opposed to non-qualitative perceptions. As a consequence, a representationalist position which does make mention of an epistemologically sound higher-order condition collapses into a view on which qualia are intrinsic properties of experiences, and hence are irreducible to intentional contents. In view of this challenge, I put forward an alternative representationalist proposal and explain how it avoids this consequence. The proposal is that qualia are global meta-representational contents.

Contents

Introduction	5
1. Towards the Representational Thesis	13
1.1 <i>Intentional Content of Experience is Narrow and so are Qualia</i>	18
1.1.1 <i>Experience Makes Us Conscious of Internal States of Affairs and so its Intentional Content is Narrow and so are Qualia</i>	19
1.1.2 <i>Experience Makes Us Conscious of External States of Affairs, but its Intentional Content is Determined Internally and so it is Narrow and so are Qualia</i>	25
1.2 <i>Intentional Content of Experience is Wide, but Qualia are Narrow</i>	47
1.2.1 <i>Functionalism and Intentional Contents</i>	49
1.2.2 <i>Against Block's Supervenience Claim</i>	56
2. Developing the Challenge	61
2.1 <i>Worldly vs. Experiential Subjectivity</i>	64
2.2 <i>Higher-Order States as Internalised Public-Language Sentences</i>	86
2.2.1 <i>Subject-Dependence of Qualia</i>	86
2.2.2 <i>Recognitional Concepts</i>	98
2.2.3 <i>HOR's Commitment to Recognitional Concepts</i>	105
2.2.4 <i>The Origin of Qualitative Concepts in Theoretical Discourse: Sellars</i>	114
2.2.5 <i>(Innate) Mentalese Concepts?</i>	134
2.3 <i>Sensa and Phenomenal Individuals</i>	142
3. An Alternative Proposal	151
3.1 <i>Qualia as Properties of Global Higher-Order States</i>	151
3.2 <i>Reviewing the Circular Argument</i>	163
3.3 <i>Higher-Order Contents Without Higher-Order States</i>	167
Conclusion	178

Acknowledgements

I am grateful to Professor Paul Noordhof for his patience and support throughout my course of study.

Introduction

“Intentionality can’t be reduced and it won’t go away”. These words sum up the manner in which Putnam distances himself in his book *Representation and Reality* from his earlier functionalist view of intentionality. Intentionality (of thoughts, beliefs, etc., as distinct from non-conceptual intentionality) is normative. The question of reducibility does not arise for normative facts – the fact that in Europe we become adults for legal purposes at the age of 18 is not to be explained in naturalistic terms. But this is not a concession to dualism, it is merely recognition of where epistemic intentionality belongs – in what Sellars calls ‘the logical Space of Reasons’. Pursuing a reductive explanation of epistemic intentionality in terms of a functionalist/computationalist model is therefore misguided. On the other hand, many feel that there must be a reductive explanation of non-conceptual intentionality – sensory perception is part of our basic phylogenetic equipment, lacks a normative dimension, and belongs in the what McDowell calls ‘the logical Space of Nature’ to contrast perception with the logical Space of Reasons. If non-conceptual intentionality can (or indeed must) be reduced on these grounds, then the suggestion that we may find immediately appealing, as I certainly do, is that qualia could perhaps be reduced together *with* it.

Representationalism about qualia offers a simple and original answer to the question as to how it is possible that we ‘see’ phenomenal colours even though nothing in the physical world including ourselves is coloured in this sense, without being dualistic or eliminativist. Seeing a phone booth across the street in, say, a three-dimensional virtual reality image by no means implies that there is an immaterial phone booth at the location where it appears to be in the image, rather, the phone booth is its topic-neutral intentional object. Similarly, the phenomenal colour of a phone booth I see in normal circumstances is not something immaterial which must exist either at the location where I see the phone booth or ‘inside’ me – it is the representational content of my experience that is neither material nor immaterial. Unlike eliminativism, treating qualia as contents is not an overreaction to dualism. Representationalism explains away rather than ignores so-called qualia intuitions on the basis of which it is argued that qualia are intrinsic and non-

relational properties of experiences. It is not surprising then that representationalism has drawn much endorsement and is nowadays a dominant position.

The work is divided into three sizeable chapters. In the opening chapter I defend externalist representationalism against objections raised on the grounds that sensory content isn't wide (and hence neither are qualia), or that wide contents and qualia come apart in certain counterfactual circumstances. The aim of the discussion is, firstly, to clarify the notion of qualitative content and locate it within the general functionalist approach of which externalist representationalism is a branch, and, secondly, to show how we can gradually progress from the view that qualia are intrinsic properties of experiences to the view that qualia are wide contents. This gradual progress towards representationalism has two main stages marked by these claims:

- (1) Intentional content of experience is narrow and so are qualia.
- (2) Intentional content of experience is wide, but qualia are narrow,

a rejection of which leads to externalist representationalism:

- (3) Intentional contents are wide and qualia are intentional contents.

In fact I begin Chapter One, for reasons I will explain below, by rehearsing Lycan's counterfactual analysis designed to dissolve phenomenal individuals (such as sense-data). (3) is established at the end of the chapter because none of the common objections (falling under (1) and (2)) to representationalism are effective.

In the core part of my work (Chapter Two) I present a challenge for representationalism from an altogether different perspective. The challenge develops from the following initial difficulty: Consider the opposite ends of the evolutionary spectrum regarding photosensitivity. At one end there are simple organisms with photosensitive tissues which have not evolved visual organs. At the other end we have, for example, humans and higher animals with far more sophisticated photosensitivity involving visual organs and, crucially, phenomenal consciousness. It is what lies in the middle of the spectrum that is the root of the initial difficulty. We may reasonably

suppose that there are organisms with visual organs that do not produce phenomenally conscious states, which organisms I call ‘unconscious perceivers’, for it is unlikely that there was a sudden evolutionary leap from simple photosensitive tissues to phenomenal consciousness – in fact any claim to the contrary would be quite incredible. If so, the representationalist owes us an explanation of why such organisms meet his conditions for attribution of wide sensory contents, since they show sufficient behavioural sensitivity to external stimulation, and yet don’t experience anything. Here I follow Carruthers’ diagnosis of this situation (although he chooses blindsight as a case of perceiving without experiencing): what these organisms share with us is worldly subjectivity, i.e. an ability to take a subjective perspective towards worldly objects, but lack experiential subjectivity, which is an ability to take a subjective perspective towards their own experiences. It follows from this that the ‘locus’ of phenomenal consciousness is a higher-order introspective state and that we must amend our representationalist characterisation of qualia so as to include a formulation of a higher-order condition to the effect that qualia are those wide contents which are (at least) available (or poised) to be picked out by a higher-order introspective state. So if at the end of Chapter One we arrived at the claim that

(i) Qualia are not intrinsic properties of perceptual states, and if so, they must be wide contents,

we now have a claim that

(ii) The thesis that qualia are wide contents can be plausibly held only as a claim of higher- rather than first-order intentionalism.

Bearing in mind the diagnosis of the initial difficulty for first-order representationalism, the difference between unconscious perceivers (or blindsighters) and us is now to be found at the level of introspective states. In other words, the move to higher-order representationalism is a step forward only if introspection directed at qualitative states features concepts which unconscious perceivers do not acquire due to

not being able to satisfy their possession conditions, otherwise higher-order representationalism cannot be thought to overcome that initial difficulty. The concepts we are after are known as phenomenal concepts, i.e. concepts which distinctively apply to phenomenal properties. Further, the concepts in question mustn't presuppose a pre-existing conceptual material supplied by public language if we don't want our higher-order position to have undesirable implications for phenomenal consciousness in very small children. One kind of concepts with the required characteristics are purely recognitional concepts. A purely recognitional concept is acquired on the basis of a brute acquaintance with a phenomenal property of certain type, its intension reduces to the causal/nomic relation to a phenomenal property.

It is at this point that I engage Sellars' reconstruction of our introspective conceptions in the discussion. He would clearly deny the existence of concepts which we come to possess in the above manner (i.e. by means of brute acquaintance) as epistemological fiction. The remainder of my work can be seen as an examination of what becomes of representationalism when it is 'filtered through' Sellars' thought regarding introspective awareness and subjectivity. I will argue in Chapter Three, where I make my alternative representationalist proposal, that representationalism is uniquely positioned to address his latter concerns to do with identifying a single and unified subject of experience. For it is open to us to replace the notion of a unified subject of experience as it is traditionally understood with the notion of a global meta-representation. But the problem for representationalism – and this is the central problem of my work – that arises from the fact that phenomenal concepts (including purely recognitional concepts) are epistemologically dubious for largely Sellarsian reasons which I'm about to explain calls for a major revision of the standard representationalist accounts. I undertake that revision in Chapter Three.

Ordinary colour terms refer to properties of external objects. It is a trivial point that these are not sensitive to spectrum inversions. Grass may appear to me the way roses appear to someone with an inverted spectrum, even though we both apply the same ordinary colour term 'green' to grass and 'red' to roses. On the other hand, a purely recognitional phenomenal concept rigidly designates a (phenomenal) property of my experience and has no natural expression in public language. Postulating concepts which

are acquired in an act of ‘brute’ recognition independently of any prior conceptualisation would be regarded by Sellars as an illegitimate attempt to prise something epistemic out of the logical Space of Reasons. So none of my concepts referring to properties of my experiences are supplied ‘privately’ from within, so to speak, they must be supplied by public-language. Consequently, my ability to focus on a property of my experience requires that I have a common-sense theory of mind, in other words, that I have a certain understanding of inner episodes, which understanding is formed in the course of learning a public language. This understanding is essentially intersubjective – inner experiential episodes are conceptualised as theoretical postulates (reported inferentially) which later become observable in the sense that they are non-inferentially reportable. In view of this rejection of phenomenal concepts, the third claim I add to (i) and (ii) goes as follows:

(iii) To have a higher-order thought about one’s own perceptual state is to token an internalised public-language sentence whose meaning is holistic and normative.

It was an insight of Stephen Leeds’ *Qualia, Awareness, Sellars* that blindsighters are not in principle precluded from developing the sort of common-sense theory of mind in which, according to Sellars, our introspective conceptions originate. They can be trained to non-inferentially report perceptual states with respect to the blind portion of their visual field in the way normally sighted people do according to Sellars. Remember that the move to higher-order representationalism was prompted by the failure to find a sufficient difference between us and unconscious perceivers at the level of first-order representation. The difficulty returns with an equal force at the level of higher-order representations. Internalised public-language sentences also fail to provide the required difference.

The first claim of my argument (claim (i)) was that qualia are extrinsic properties of experiences and if so, they must be wide contents. But now it seems that we cannot distinguish sighted creatures from unconscious perceivers in terms of wide contents even if we add the aforementioned higher-order condition, namely, that qualia are those wide contents that are available to be picked out by higher-order meta-representations. The immediate consequence of this is that qualia are, after all, intrinsic – our states have them,

whereas the states of creatures who perceive without experiencing don't. Hence the following claim:

(iv) Since higher-order perceptual thoughts are internalised public-language sentences, qualia are intrinsic properties of experiences.

Accepting that qualia are intrinsic is a substantial concession to dualism. Claim (iv) clearly contradicts claim (i) where qualia are said to be extrinsic, which brings us to the central problem in my work: I said at the very beginning that the question of reducibility does not arise for epistemic intentionality precisely because it belongs in the logical Space of Reasons and that, on the other hand, we should rightly expect phenomenal consciousness to be reducible because it belongs in the logical Space of Nature. However, if we build epistemic intentionality (i.e. the intentionality of introspective states understood as internalised public-language sentences) into the architecture of phenomenal consciousness as we did above, and the resulting representationalist position collapses into a view that qualia are intrinsic, a naturalistic explanation of phenomenal consciousness in representationalist (or indeed any other) terms will prove elusive.

Chapter Three is an attempt to put forward an alternative higher-order representationalist view which escapes that consequence. Firstly, I suggest that we think of the relevant higher-order states as non-epistemic in order not to rely on phenomenal concepts to provide the difference between us and blindsighters. It may be objected that this move could have been made earlier and that I should not have restricted my discussion of higher-order states to thoughts, so the central problem would not have arisen. There are of course Inner Sense theories of introspection, but on the Sellarsian approach I adopt introspection presupposes a common-sense theory of mind developed in the way briefly described above. Introspection is therefore a conceptual capacity and Sellars would, again, dismiss Inner Sense theories of introspection as an attempt to force something epistemic outside the logical Space of Reasons. Bearing these points in mind, my suggestion is not only that the higher-order states are non-epistemic, but also that they are not introspective. Introspection is an independent capacity acquired later in life, while non-epistemic meta-representations that are built into the higher-order structure of

phenomenal consciousness continuously target lower-order contents and are available from birth. Another notable advantage of this is that we need not deny phenomenal consciousness to pre-linguistic children. Such a commitment to non-epistemic meta-representations also involves shifting qualia to higher-order level in that higher-order contents are qualitative but lower-order contents are not. Suppose, for illustration, that I am now in a situation where there is a non-qualitative non-epistemic meta-representation aimed at a lower-order quale. Since my awareness of the lower-order quale is facilitated by the meta-representation, whose content has no phenomenal character, I would not experience anything in these circumstances.

As I present the central argument in Chapter Two, another important problem regarding phenomenal consciousness comes into focus, one which representationalism can deal with in its own specific way. To address the problem, I suggest that it is in fact not a multitude of individual meta-representations that target lower-order states, rather, there is a single global meta-representation which is continuously directed at all lower-order perceptual and as well as proprioceptive contents. Sellars does acknowledge the problem near the end of his essay *Phenomenalism* where he says that sense impressions are attributable to a single and unified self (or subject), although there is no such self to be found. We are made up of micro-physical parts none of which serve as subjects to which sense impressions are attributable. A (non-representationalist) functionalist model will not solve the difficulty either, for it consists of a number of interacting functionally described components (modules, sub-routines, etc.), each of which is a sub-personal part and, again, cannot serve a subject to which sense impressions are attributable. It is partly for this reason that he turns, as he puts it, each sensing into a logical subject in its own right and introduces *sensa* (or phantasms) as the Scientific Image replacement for sense impressions, whose home framework is the Manifest Image. I argue that the move to *sensa* can be avoided by introducing the idea of a global (qualitative) meta-representational content; the meta-representation itself is sub-personal, as it is presumably a brain state, but its content, due to its global scope, is a personal level one. Global meta-representational content is introduced as a replacement for a single and unified subject as it is traditionally understood. Finally, the existence of a global meta-

representation is certainly consistent with the fact that we are made up of micro-physical parts, since at least those in our brains can be thought of as realising the global state.

1. Towards the Representational Thesis

Sensory experience is analysed into these three components: it has an intentional content, a functional role and a qualitative aspect. The intentional content of an experience expresses what the experience is about, or the properties of what object we are aware of in our undergoing a certain experience. Attributing intentional content to experiences carries the assumption that experiences have certain representational properties. The experience's functional role, on the other hand, relates the experience to its sensory input, other mental states and motor outputs (or more generally, behavioural outputs).

The qualitative aspect, or qualia, is the only component of experience that is purported to be non-relational or intrinsic to the experience and can therefore be reduced to neither the intentional content of the experience nor its functional role. The term 'qualia' is used here in its more rigid sense, meaning that if one talks of experience in relational terms he necessarily leaves out qualia. In the looser sense the term 'quale' is interchangeable with 'what-it's-likeness' or 'raw feel', without further commenting on whether this is relational or not. If some authors maintain that there exist no such things as qualia, it only amounts to rejecting that there are qualia as defined in the former way, and not to claiming that there is nothing it is like for us to undergo an experience. For illustration purposes I will use the term in its strict sense for a moment and I will notify the reader of changes made afterwards. In this rather highly theoretical sense of the word then, it is by no means obvious how a correlate of it is to be traced in ordinary language or common-sense psychology. In perceiving, our qualitative property ascriptions seem to regard external objects (it is hardly common practice among us as ordinary folk to describe our perceptual states as being pink or triangular), although it may be pointed out that in certain non-standard conditions the converse is demonstrated by thoughts such as "this looks to be blue, although it is in fact purple", when we allegedly direct our attention at certain features of the experience itself - by having such a thought I come to distinguish between the property of the perceived object on the one hand, and what my experience presents the object as on the other and thus, some would urge, the (qualitative) properties of my experience. Undeniably, such scenarios do occasionally induce in us higher-order (or introspective) beliefs about our own perceptual states. At the same time, though,

common-sense psychology provides no further clues as to whether “I have a pink and triangular experience” does not simply stand for “I’m in a state that presents me with a pink and triangular object”. Ordinary language does little to reinstate qualia.

One pre-theoretic temptation to insist that higher-order beliefs of the above variety are about the properties of experiences rather than their intentional objects rests on likening perceptual states to pictures. To see how embracing this pre-theoretic association with pictures uncritically leads one to believe in qualia, consider this illustration: Suppose that a cinema screen depicting an oncoming train is an analogue of a visual experience of an oncoming train. Now imagine that one person in the audience doesn’t understand how screens generally work, he hasn’t seen one before and no one has explained to him that what he is in fact viewing is a two-dimensional projection of light rays of various colours of the spectrum on to a piece of canvas. He thus believes he is in danger of being run over by an actual train. Another person, by some strange misfortune, is unable to interpret what he is viewing as a depiction of a moving object, but in all other respects he is just like us. He is also aware that he has a flat piece of canvas in front of him. To complete the story, the rest of the audience suffers no such impairments and understands perfectly well that there is a depiction of an approaching object in front of them. If we now apply the above definition of the distinction between intentional contents and qualia, it follows that in this example the person who is sincerely frightened by what he sees is aware merely of the representational content of the moving picture due to his failure to appreciate the intrinsic features of the picture, which is in turn an inability to comprehend that the representational content (of the picture) he is aware of is in this case given by the spatial relations obtaining among two-dimensional areas of different colours and shapes on a flat surface presented to him.¹ The situation in which the other person finds himself is quite the opposite. He is unaware of the representational content (as all he sees is a rapidly expanding patch of colour, without the third dimension), but is well aware of the intrinsic features of the picture. If we strain the terminology a little, we may conclude that

¹ Michael Tye argues (1992) that pictures are not analogous to visual experiences in that it is disputable whether experiences have their contents determined by their qualitative aspects. Indeed, in this article he rejects that there are qualia in this sense. But more on this further below. The purpose of my present example is to bring out how some philosophers make sense of the idea that qualia are intrinsic to experiences.

the first person is not aware of the ‘quale’ of the picture, whereas the second one is (the rest of the audience will be capable of shifting focus from one to the other).²

The case of the person who is only aware of the intrinsic features of the picture models the aforementioned pre-theoretic conception of how introspective (higher-order) beliefs about our perceptual states can uncover their intrinsic qualitative features. If the model goes unquestioned, it is easy to see why many are adamant that the notions of intentional content and functional role on the one hand, and the notion of qualitative aspect on the other must be *strictly* kept apart. For the model will simultaneously be used to articulate our awareness of the intentional content of our experiences (this, in the example, is the case of the person who was sincerely frightened of what the screen confronted him with) – such awareness will be mediated by an analogue of a picture with non-relational qualitative features. Qualia, it would be concluded, can’t be reduced to intentional contents – qualia *explain* how they arise (they are their source – just as in the example the rapidly expanding patch explained how one could see an approaching train), and therefore precede intentional contents in order of explanation. In other words, something along these lines may lie at the bottom of the widespread reluctance to define qualia in the looser sense. From now on I will use the term ‘quale’ in its looser sense, which is merely for the sake of convenience as it will spare me making verbal points almost each time the term is used. ‘Quale’ will not be suggestive of whether mental qualities are intrinsic or relational in nature. The following considerations will be those of the status of qualia rather than of their existence (see footnote 1).

The fully developed philosophical positions that are compatible with and lend theoretical support to the preconception just discussed take on a variety of forms. Their unifying feature is, in accordance with the preconception, the way they view the relation between intentional content and qualia, namely that the latter determine the former. In this chapter I set myself the task of proceeding towards a reversal of this determination relation and thereby arriving at the representational thesis (as formulated in (3) below). Although pursuing the task will inevitably involve dismantling all fundamentals of the

² Note, again, that this is only an illustration and I do not mean to attribute qualia to pictures. Qualia, however they are construed, are mental features. Nor do I wish to suggest that there is nothing it is like for the person who is unaware of the ‘quale’ of the picture to see the screen. The screen itself will be the object of his experience (and not the train) producing its own quale in him and therefore there will be something it is like for him.

theories that are in the spirit of the preconception, having done so will not by itself complete the reversal. In fact there will be a number of stages to it - we will, on behalf of the representationalist, advance through a terrain of philosophical theories surveyed and mapped in terms of where they stand on the issue of narrowness vs. wideness of content as well as narrowness vs. wideness of qualia. One highly desirable by-product of this strategy will be a fairly detailed characterisation of representationalism – its essential characteristics, scope, argumentative resources and explanatory power will gradually come to surface in the course of the reversal.

In view of the task thus laid out and the way I intend to go about it, I will present representationalism as a thesis one arrives at once one has successfully refuted the claim that intentional contents of experiences are narrow and qualia are narrow and then the claim that although intentional contents are wide, qualia are narrow (the two conceptions will be the two major stages in the process of the reversal). Therefore, these assertions will be the subject of my discussion:

- (1) Intentional content of experience is narrow and so are qualia.
- (2) Intentional content of experience is wide, but qualia are narrow.

, a successful refutation of which will leave us with the following:

- (3) Intentional contents are wide and qualia are intentional contents (representationalism).

Some general comments on the scope of (3) are immediately in place. Plainly, the mention of ‘wide’ in (3) takes pains outside of its scope. This does not by itself mean that a suitable representational account of bodily sensations cannot be given, despite their inevitably being only narrowly intentional. Those who are already sceptical about representationalism as a theory of perception will generally find a representational treatment of pains even less palatable – while few doubt that perceptions have intentional contents, it remains under dispute whether bodily sensations exhibit any sort of intentionality at all. But representationalism about bodily sensations does not always come out worse off than representationalism about perceptions. Take, for instance,

Block's (1990) rejection of the theory in general on the grounds that qualia supervene on physical (to mean physicochemical or molecular) constitution (although I shall argue extensively in section 1.2. that Block is wrong). Whereas the supervenience claim clearly clashes with the former (if true, perceptual qualia can't even be narrowly intentional), its truth is insufficient to establish that pain qualia aren't intentional. To appreciate this, consider the representationalist's perspective on the issue of pains: my pain or ache is about a discomfort felt at a particular location in or on the surface of my body, making me wish to rid myself of whatever is causing it. It also alerts me to (and so is also about) which type of tissue is affected, not least because I can assess the risk to my health that may result from the disturbance of that particular type of tissue. Therefore, having my earlobe pierced will feel differently to having my palm pierced (with, say, the same needle) due to them, physicochemically speaking, being different types of tissue. Now, given that we can feel either a pain or an ache in virtually every part of our body, for two individuals' pains and aches to have the same intentional contents it is perhaps necessary that their bodies be made up of the same variety of types of tissue and hence that they be physically identical. The case for intentionality of pains (endorsed already by Pitcher (1970) and Armstrong (1962)) can be pressed alongside the supervenience claim.

My second remark on the scope of (3) concerns qualitative differences among states belonging to distinct sense modalities. While there was no question about where pains stand in relation to the scope of (3), here, opinion among qualia externalists themselves is divided. Lycan (1996) suspends (3) as a possible explanation of this kind of qualitative difference and turns instead to differences in proximal stimulation for more promising results. One familiar example that would appear to point towards such differences falling outside the scope of (3) sets a visual sensation of a smooth surface against a tactile sensation of a smooth surface: same wide content, distinct qualities. The immediate answer is that much here hinges on how content is made explicit, for we have as yet no independent grounds to hold that the two states' contents should not turn out to be distinct. I propose that a final verdict on the matter be postponed until it is put under closer scrutiny. However, the character of some of the discussions that follow in this chapter will create suitable setting for these questions to be brought up as we go along, giving increasingly broad hints as to where the answer should lie.

I will now embark on the reversal announced in the opening paragraphs. The next section will carry us through its initial stage, that is, a move from (1) to (2).

1.1 Intentional Content of Experience is Narrow and so are Qualia

We already touched on the topic of narrow content in connection with bodily sensations. This only covers one sense in which content can be said to be narrow. There are others. For the wide/narrow distinction with respect to intentional contents is not understood as being parallel with the external/internal distinction with respect to the objects that we are aware of. To spell this out, even if a theory has it that, in experiencing, we are aware of external states of affairs, it can not be concluded that it *necessarily* supports the alternative that intentional contents of experiences are wide. On the other hand, if a theory has it that experiences make us aware of internal objects – whether they be mental (e.g. sense-data) or physical (disturbed tissue, as in the case of pains) it can be safely concluded that it amounts to claiming that intentional contents are narrow. Intentional content of experience is narrow either if it is the case that experience makes us conscious of mental objects (those existing in the mind), or if it is the case that although experience makes us conscious of external objects, the intentional content of experience is determined *internally*. (1) thus allows these two alternatives:

(1a) Experience makes us conscious of internal states of affairs and so its intentional content is narrow and so are qualia.

(1b) Experience makes us conscious of external states of affairs, but its intentional content is determined internally and so it is narrow and so are qualia.

It also helps to think of the wide/narrow distinction as applied to intentional contents in terms of supervenience: if intentional contents of experience supervene on causal, contextual relations, they are wide, and if they do not co-vary with variations in environmental factors, they are narrow. This should make obvious the reason for which in (1b) intentional content is still narrow. (1b) does not make the causal links relevant in

the requisite way³. Distinguishing (1a) and (1b) further fragments the reversal I have undertaken to complete, so the move from (1) to (2) will in fact be a move from (1a) to (1b) and then from (1b) to (2). I now turn to (1a).

1.1.1 Experience Makes Us Conscious of Internal States of Affairs and so its Intentional Content is Narrow and so are Qualia

According to (1a), our experiences are immediately about objects existing in the realm of the mental and if they refer to external objects, they do so in a derived way. For instance, in hallucinating a mermaid, although there is nothing in front of my eyes, I am certainly conscious of something and this something must exist in my head. The experience shares all its introspectible properties with one I would undergo were a real mermaid present (if there were such things as mermaids) save for the fact that its object does not actually take up a region of space (we may say that it is temporal only). After-images are supposed to be similar cases to the one just mentioned. They make us conscious of a mental object which becomes the intentional object of the sensation. The fact that the very same sensation-type sometimes happens to be causally correlated with an external object is accounted for by saying that the causation is indirect and that the external object only causes a mental object to occur in the mind which is then sensed. The argument that is very often alleged to sustain this approach states that the perceived properties are in many respects unlike the properties we ascribe to physical objects. The argument also entails that the properties we associate with physical objects are the result

³ We should be wary of possible misinterpretations of the co-variation. For instance, John Searle (1983) maintains that the same (narrow) psychological states occurring in two different subjects can nonetheless determine distinct intentional contents, which may, at a first glance, seem to be consistent with saying that intentional contents co-vary with changes in environmental factors. The appearance of consistency stems from the fact that on this view, by thinking “Water is wet” on Twin Earth Twin Oscar will refer to XYZ (XYZ is the chemical composition of the liquid that fills rivers and lakes on Twin Earth, as opposed to H₂O on Earth) and not to H₂O. But the external factor here – XYZ – is not meant to be the reason why Twin Oscar’s thought has a different content from that of Oscar’s thought on Earth. It is because of something internal that Twin Oscar’s thought has a different content.

On the other hand, representationalists have come up with a teleological response to Block’s version of the thought experiment involving visual perception. The response is perfectly in accordance with the co-variation thesis, although the contents of the traveller’s experiences will not shift on his arrival at Block’s Twin Earth (the experiment as well as the responses will be examined later).

of the process of learning how to derive them from the perceived ones. Take, for example, the property of something being parallel with something else. If I stand on a railway track, staring at the two lines as they reach up to the horizon I perceive them as convergent. It is solely by means of reasoning and other cognitive faculties we possess that from the perceived convergence we are able to infer to the objects' real features. If I walk to the point where the horizon seemed to be from the spot where I was previously standing and then look back, I won't see any divergent lines, but again, there will be two lines converging at just the same rate as before, from which I will figure out, having enough similar experience with other objects, that they are parallel.

This apparently strikes a chord with the pictorial model of sensory perception laid bare in the illustration at the beginning of this chapter that was seen as the precise opposite of representationalism, as being virtually its mirror image, embodying all that the representationalist will want to distance himself from. To get ourselves off the ground and begin progressing towards (3) we must expel the kind of mental object from the mind that the model implants there to mediate between the world and our perceptual awareness of it. Let's focus on just one (vivid) example of such ostensible mental objects - after-images. The choice isn't arbitrary. Traditional approaches like (1a) often extrapolate from non-veridical cases to veridical ones, not otherwise, since non-veridical states are often held up as supreme evidence of there being mental targets of perceiving (to repeat what was said above, in having a green after-image, there is nothing green out there, so it must be in my mind). Showing that there are no such things in non-veridical cases will remove the source of the positing of the mediators allegedly involved in veridical ones.

One strategy for avoiding talk of mental objects (or sense-data talk) that instantly springs to mind is to introduce adverbial translations for sentences like "I am sensing a green patch" (uttered in situations when I have a green after-image) which, without such translations, clearly entail that "there is a green patch that I am sensing". Adverbialism was designed to sever this entailment link through which the sense-data theory gains its force. If topic-neutrality proves unattainable (whether this is done adverbially or by some other means), the object involved in the quantification will necessarily be mental, since there is certainly no green physical object in my brain when I have a green after-image. Michael Tye (1995), for example, mentions two types of theories of adverbs: an event

theory and a predicate operator theory. According to the former, a sentence like “I am sensing greenly” is analysed as “There is an event of sensing that has me as its subject, and that event is green”. Tye says that one of the difficulties here, among others, is that ‘that event is green’ is unintelligible, although quantification over events should not itself be a problem considering the obvious advantages they bring in contrast with mental objects. The latter adverbial theory does succeed in not introducing entities except the subject itself, on which, he says, statements about mental objects are in fact statements about sentient creatures and the phenomenal properties they instantiate. Nevertheless, with example sentences like “I sense quadruply-pinkly”, which is the predicate operator translation of “I have four pink after-images”, we are incapable of making the apparently valid inferences to sentences like “I have fewer than seven after-images”.

So can we propose a type of topic-neutral translations that do not quantify over mental objects? I will follow Lycan’s investigations in (1987, p.85-89) for a moment to demonstrate that there are topic-neutral translations that not only overcome all the difficulties adverbialism normally runs into, but also that there is much in the outcome of this analysis that the representationalist will wish to build and elaborate on. I’ll try to justify the latter claim as soon as I’ve shown how Lycan’s discussion proceeds. Lycan begins his considerations by clarifying the syntax of “Leopold is sensing homogeneously-greenly”, homogeneity being a property of having the same colour throughout. He observes that homogeneity cannot be ascribed to the *property* greenness, which is what ‘homogeneously-greenly’ is partly made up of. To make homogeneity a first-order property (a property of an object rather than just another property), and in order for the sentence to retain its adverbial form he gives this alternative: “Leopold is sensing a-homogeneously-green-patch-ly”. ‘Homogeneous’ is now ascribed to a green patch instead of greenness itself. However, a different problem arises. Lycan asks how it is possible to understand such a complex predicate and, worse, how it is possible to understand a predicate containing a noun, namely, ‘green patch’, generally. What is its semantics? He urges us to understand it as an instrumental adverb, that is, along the lines of “Leopold is buttering with-a-knife-ly”, which is an adverbial translation of “Leopold is buttering with a knife”. On this construal, the sentence will entail that “There is a green patch that Leopold is sensing”, just as “Leopold is buttering-with-a-knife-ly” entails that

“There is a knife with which Leopold is buttering”. To take this route would be to favour sense-data talk again. We have seen, on the other hand, that he cannot dispense with including some grammatical form of ‘green patch’ in the sentence, because he needs to keep ‘homogeneity’ a first-order property. So the only way out of the trouble, he suggests, is to look for an instrumental adverb that does not bring the presence of an actual object with itself. This can be achieved by rephrasing adverbial sentences as counterfactuals. Lycan gives us another example: “Leopold is running as if a lion were chasing him”. It is a counterfactual rendering of “Leopold is running a-lion-is-chasing-him-ly” and in this case it can easily be argued that the sentence does not require that there be a lion chasing Leopold. It does not entail that “There is a lion chasing Leopold”. When applied to sensing, the reasoning leads to the following result: “Leopold is sensing as he would be sensing if a green patch were present to him”. There is nothing in the sentence that suggests that there is an (actual) green patch in front of Leopold. Hence, the question as to whether what Leopold is seeing is mental or physical simply becomes irrelevant. We have an uncontroversial example of topic-neutrality. It fits into this general model:

There is a mode of sensing M such that Leopold is sensing in way M and Leopold would be sensing in way M if a green patch were present to him and conditions were normal,

and Lycan adds that

the antecedent of this counterfactual is “there is a green patch present to Leopold”. Thus, any of the standard accounts mentioned above directs us to look at an alternative world in which (it is really true that) there is a green patch present to Leopold. And this is the source of our quantifier: the value of its variable *is* a green – physically green – patch (p.87-88, italics original).

According to Lycan, having explained after-images in this way we can now specify the truth-conditions for “Leopold is sensing a-homogeneously-green-patch-ly” by an

appeal to whichever particular semantics for counterfactuals we prefer. We are also able to determine the meaning of the slightly overcomplicated adverb ‘a-homogeneously-green-patch-ly’, because the puzzling ‘green-patch’, by being in fact a part of a counterfactual, will refer to an object existing at a possible world.

Let us now take stock of what we have arrived at. On (1a), what we directly perceive in both veridical (e.g. a sense-datum presenting me with convergent lines, as in the example above) and non-veridical cases (e.g. an after-image) are (ontologically) mental objects. Clearly, if we allow this, the states that constitute awareness of such objects must also be ontologically mental, since it is out of the question that something immaterial should be sensed by something material. The counterfactual treatment removes mental objects of perception from the mind, hence the states that constitute perceptual awareness being ontologically mental is no longer a matter of necessity (although it is of course the translation itself that guarantees this in the first place). We can now conclude that in fact two types of topic-neutrality emerge from this way of purging the mind of mental objects of perception: that with respect to *what* is represented and that with respect to the state that does the representing.

Seeing that the representationalist will wish to embrace both is revelatory of what type of position representationalism is. The counterfactual analysis implies that what is to be found in the mind is only a perceptual mental state (in which an object is represented, whether existent or non-existent), not the perceived object. The mental representation may give rise in us to thoughts such as “There is a green patch in front of me”, but this is as reliable evidence of the presence of an object as someone’s reporting that there is a cat in front him while looking at an extremely realistic painting. All he is physically presented with is a canvas and paint arranged in such a way that it happens to represent a cat. No one, having been told what he is in fact staring at, would believe that some non-spatial cat must nonetheless somehow be there. This is, in principle, how topic-neutrality of *what* is represented should be understood. In principle only, because as far as representationalism is concerned perceptions do not represent in the same manner as pictures do. One of the reasons for this is that representationalists are typically materialists, which means that mental representations are supposed to be realized in the brain-matter. Neurons do not constitute pictures in any literal sense. Tye (1991), for

instance, favours the hypothesis that veridical as well as non-veridical perceptions share the same mechanism taking place in our visual cortex consisting of matrices of symbol-filled cells, with symbols encoding information about various properties of physical objects. But the symbols themselves are not coloured. The encoding function they perform is (chemically?) hardwired. It is easy to construct a counterfactual of the above type for the case of after-images if this particular hypothesis is adopted to explain mental representations: The matrices of cells in the visual cortex of subject S who is having a green after-image are filled by the same symbols as they would be if a real green patch were present. Non-veridical perceptions are situations which cause our visual system to react in the same way as it would react if the perception was veridical. To put it rather loosely, when for some reason our brain changes its state to that in which it seems to us that there is something green in front of us, we can do nothing but perceive it, no matter whether we do or do not believe what it presents us with.

Having said this, it is important to emphasize again that the fact that representationalists are typically materialists is not the crucial reason why they commonly deny that perceptions represent in the same way as pictures. We could quite plausibly imagine a representational theory of perception on which representations would not be realised by the brain. This explains why the theory is also topic-neutral with respect to the states that do the representing. Dualism is not a priori incompatible with representationalism. A creature consisting of both body and soul could still be a representational system, with representations taking place in the soul. Thus, the issue of what realises representations is more an empirical one and cannot be resolved on a priori grounds⁴. The representationalist can leave it open whether representations are ontologically mental or not. What he does not leave open is whether they represent like pictures. They do not, it is claimed, because experiences do not have any intrinsic phenomenal properties. Or, at least, there are no intrinsic phenomenal properties we are conscious of in undergoing an experience.⁵ If, after enough empirical research has been carried out, representations turn out to be functional states of the brain (as representationalists typically assume) then the relevant physical-chemical properties will

⁴ The same is true of functionalism in general. See, for example, (Putnam 1967).

⁵ For this alternative see (Harman 1989).

indeed be intrinsic to them. Representationalism and the more traditional views have things the opposite way round: representationalists hold that it is impossible to account for the what-it's-likeness without referring to (wide) intentional contents, whereas on the traditional view the intrinsic phenomenal properties (if there are such things) do not only explain the what-it's-likeness of experience, but in many cases they are also supposed to determine intentional contents.

I hope these remarks have helped clarify why I believe that Lycan's counterfactual analysis provides suitable ground for the representationalist to build on. However, there is nothing in the analysis to stipulate that that ground is to be occupied solely by an externalist theory (of qualia). It does link Leopold's state to a physically green object, but it falls short of showing that this physical greenness is what gives the state its qualitative identity. The analysis can equally well be complemented by any of the instances of (1b), which, despite not implanting any ontologically mental mediators between the world and the mind's perceptual awareness of it, remain internalist in that it is entirely features of the subject that impart qualitative identity to his experiences – the instances differ merely over which particular features of the subject these should be. We nonetheless conclude this sub-section content that the important initial step in the direction of (3) has been made. I now move on to (1b).

1.1.2 Experience Makes Us Conscious of External States of Affairs, but its Intentional Content is Determined Internally and so it is Narrow and so are Qualia

There are at least three kinds of views that ought not to be overlooked in connection with (1b). They all fit into (1b), although each does so for slightly different reasons. I'll begin by examining with John Searle's treatment of experiential content in his (1983). The underlying assumption that constrains Searle's thought is this: experiences have intentional contents, because they have conditions of satisfaction, and the contents are propositional, which is "an immediate consequence of the fact that they have conditions of satisfaction, for conditions of satisfaction are always that such and such is the case" (p. 41). In other words, Searle conjectures that if experiences are intentional states, they have

conditions of satisfaction and if they have conditions of satisfaction, their content is propositional. This means that experiential content is made explicit by “x sees that such and such is the case” rather than by “x sees y”. The former alternative being intensional, he says, it warrants a smaller set of inferences than the latter one (the latter one being extensional). For instance, “Seeing that Mr. Cameron is giving a speech” does not imply “Seeing that the youngest leader of the conservative party is giving a speech”, even though Mr. Cameron is in fact the youngest leader of the conservative party, whereas “seeing Mr. Cameron give a speech” does imply “seeing the youngest leader of the conservative party give a speech”, as well as other facts true of Mr. Cameron included in the “x sees y” form. To Searle, such distinction between intensionality and extensionality parallels the distinction between intentional content and intentional object. Seeing what he takes the relation between the two to be will be crucial for our understanding of why his views of experiential content fall under (1b).

The following fragment of Searle’s text will take us straight to the point we are after. There he specifies the truth conditions of “X sees that there is a yellow station wagon in front of X”, which proposition is the content of an experience of a yellow station wagon.

1. X has a visual experience which has:
 - a. certain conditions of satisfaction
 - b. certain phenomenal properties.
2. The conditions of satisfaction are: that there is a yellow station wagon in front of X and the fact that there is a station wagon in front of X is causing the visual experience.
3. The phenomenal properties are such as to determine that the conditions of satisfaction are as described in 2. That is, those conditions of satisfaction are determined by the experience.
4. The form of the causal relation in the conditions of satisfaction is continuous and regular Intentional causation.
5. The conditions of satisfaction are in fact satisfied. That is, there actually is a station wagon causing (in the manner described in 4) the visual experience (described in 3) which has the intentional content (described in 2) (p. 61-62).

What 3. suggests is that intentional content is determined phenomenally. At a first glance, this conclusion may not be so obvious, as 3. speaks of determination of conditions of satisfaction, rather than of intentional content. But they are the conditions of satisfaction of the above proposition which is the intentional content of the experience in question. Hence I think that such interpretation is perfectly fair and it is indeed at the experiential level alone that intentional content is specified. Nothing in Searle's account requires that we look outside the experience itself (that we look at what causes it, for instance) in order to specify its content (together with its conditions of satisfaction).

If this sounds unconvincing, consider also what Searle says earlier on p. 48 about how the intentional content of experience is to be made explicit: "I have a visual experience that there is a yellow station wagon there and that there is a yellow station wagon there is causing this visual experience", which he then spells out by saying that "the intentional content of the visual experience determines under what conditions it is satisfied or not satisfied, what must be the case in order that it be, as they say, 'veridical'" (p.48). Content is self-referential, meaning that it makes reference to the object which is the standard cause of the experience. Now, although points 2. and 4. in the above quotation do incorporate causality into Searle's account, causality is not intended to be prior to the phenomenal properties of the experience when it comes to determining content. To sum up, Searle's proposal is that phenomenal properties specify content and content (thus specified) only subsequently specifies the standard cause. This order of determination makes Searle's position straightforwardly an instance of (1b). Having outlined Searle's position, the task is to assess how powerful it really is as an objection to the claim that intentional contents are wide. In so doing I shall also attempt to overcome the objection and move on to (2), just as in the previous section I tried to provide reasons to abandon (1a) and move at least to (1b).

Michael Devitt's (1990) has a very similar objective to the one pursued here, although his primary concern is with the issue of thought content. In that article he highlights the contrast between Searle's view and externalism with respect to the importance they attach to the causal links by saying that on Searle's view, *content* determines object, whereas

externalism has it that *object* determines content. However succinct a way to put it this is, it may encourage a kind of misapprehension that (for reasons I will explain in just a moment) must be avoided here at all costs. What we must be cautious about is that, in the claim that object determines content, we do not understand the term ‘content’ in the same way as Searle does (this of course does not apply to the claim that the content determines the object, as it is Searle’s own claim!). On page 16. I said that Searle introduces the distinction between content and object as a distinction between proposition, which is, according to him, *intensional* with respect to the possibility of substitution (‘x sees that such and such is the case’) and *extension*, which is the intentional object stripped of all modes of presentation (i.e. the y in ‘x sees y’). Bearing this in mind, the point is that if Devitt’s formulation of externalism (i.e. that object determines content) is read as “extension determines intension”, the result is something which no externalist has ever wished to defend.

The oddity of attributing such claim to externalism is best exposed by drawing attention to thought contents, for which the application of the distinction between intension and extension is much less problematic. I do not think anyone is willing to believe that twin-earth experiments and the notion of rigid designation have been put forward to show that extension determines intension. Intension is the mode under which a referent is presented, or, the way we come to understand the associated term (via its intension). Understanding meaning is undoubtedly a psychological state. Putnam (1975, 1981) and Kripke (1972) argued that meaning, as opposed to *understanding* meaning (i.e. intension and/or the relevant psychological state), is a non-psychological matter, because it is individuated causally. The character in Putnam’s story uses the term ‘water’ when the chemical constitution of water has yet to be discovered. The narrow psychological state he is in when using the term can be described as ‘a transparent, odourless liquid that fills rivers and lakes’. Putnam says that his thoughts containing ‘water’ are nonetheless about H₂O. He imagines that the natives of a different planet also use the term ‘water’, which they understand as ‘a transparent, odourless liquid that fills rivers and lakes’, even though the chemical composition of water there is XYZ instead of H₂O. Thus thoughts containing ‘water’ on Earth are about H₂O, while thoughts containing ‘water’ on the other planet are about XYZ. The upshot is, briefly, that ‘a transparent, odourless liquid

that fills rivers and lakes' fails to differentiate between thoughts about H₂O and those about XYZ and therefore cannot be regarded as constitutive of content. The purpose of the experiment surely is not to demonstrate that the causal factor (extension) determines the intension (i.e. the narrow psychological state).

Although I have been using 'meaning' and 'content' as if they are interchangeable, I do not mean to conflate them, as meaning is of course a language-world relation, while intentionality is a mind-world relation. However, the idea I would like to make clear does not depend on the difference. Let us return to Searle and experiential content. What I will now hope to make obvious is that Searle's dismissal of the causal theories of experiential content is based on the mistake described above. His mistake is, in other words, that he first imposes his own construal of content on the formulation of externalism (by 'formulation of externalism' I understand Devitt's "object determines content") and only then he gives reasons for rejecting it. So what he in fact rejects is the claim that "object determines intensional-propositional content", even though, as I explained at some length in the previous two paragraphs, no externalist wants to say this. But it is time I provided some textual evidence.

Searle raises his objections to the causal theory of experiential content when explaining how his own conception can deal with the particularity problem. The particularity problem is a problem of how two phenomenally identical states caused by two identical (but numerically different) objects can have different contents. I think many would be inclined to believe that externalism is better equipped to cope with the problem. However, Searle clearly does not share the inclination. He describes a version of Putnam's twin earth. Our earth Jones sees his wife, Sally, getting out of their station wagon and on twin earth, twin Jones sees twin Sally getting out of their station wagon. The externalist's solution to this would be that Jones' experience has Sally as its content because Sally is the standard cause of his experience and twin Jones' experience has twin Sally as its content because twin Sally is the standard cause of his experience. But Searle is not convinced. He says: "... the problem as I have posed it is a first-person internal problem" (p.63). And his question is:

... what is it about the *content* of Jones' visual experience that makes the presence of Sally rather than twin Sally part of the conditions of satisfaction of his visual experience, and what makes the presence of twin Sally part of the conditions of satisfaction of twin Jones' experience? (p. 62, italics original)

What he understands by the italicised 'content' in the quotation is intensional-propositional content (that is, again, "x sees that such and such is the case" as opposed to extension – the y in "x sees y"), which is apparent from the fact that he takes the problem in question to be a first-person *internal* problem. I pointed out earlier that intensionality and first-person point of view are closely connected in the sense that in semantics, for example, intension is a psychological state or the way one understands the meaning of a term. So what Searle sets out to investigate is how a sensory state with an intensional-propositional content can be about Sally, rather than twin Sally. My main objection is that the stipulation that content must be intensional-propositional is too restrictive with respect to what kind of answer Searle will consider acceptable. His dissatisfaction with the causal response stems from the fact that it does not explain how causal factors, which are extrinsic to the experience, enter into the intensional-propositional intentional content and thus does not explain how these third-person facts make a difference from first-person point of view. But, as I stressed above, externalism does not purport to succeed in doing this. What Putnam attempted was not to show that the fact that thoughts containing 'water' on earth are about H₂O rather than XYZ makes a difference to how Oscar understands the term. He still understands it as 'transparent, odourless liquid that fills rivers and lakes', which is exactly the same as how twin Oscar understands it. On the externalist's proposal, these extrinsic facts do not *enter* into the intensional-propositional content, they are *constitutive* of content, which means that, if true, content is not what Searle takes it to be, it is not intensional. Searle and the externalist work with different notions of content, which I think Searle tends to overlook. Therefore, he first merely assumes that his own construal of content is correct and only then rejects externalism.

What is Searle's solution to the particularity problem? We learn that intentional contents do not determine their conditions of satisfaction in isolation. What enters into

the content of his experience of Sally are the contents of a set of past experiences caused by the same woman, which are now memory contents. Thus the content of Jones' present experience of Sally is that a woman with identical Sally-like features is before me and her presence and features are causing this visual experience and that woman is identical with the woman whose presence and features caused his past experiences of a woman whom he has known as Sally. Searle adds that if Jones is transported to twin earth the content expressed in the last sentence is not satisfied because the woman he now has an experience of (twin Sally) is not identical with the woman whose presence and features caused his past experiences of a woman whom he has known as Sally. If Jones is transported to twin earth at birth then the content is satisfied because the content of his present experience matches the contents of his past experiences (he has only had experiences of twin Sally). Such outcome implies that even though experiential content is determined internally on Searle's view, its conditions of satisfaction are particular rather than general. He calls the set of contents of the relevant past experiences 'the network'. Note that facts about the network are extrinsic to the present experience, however, they are not external facts and hence Searle's appeal to the network should not make us doubt that he is a proponent of (1b).

Given that facts about the network and facts about causal factors are both extrinsic to the experience, one may wonder for a moment why Searle is willing to allow the former to get into the experience's content but not the latter. Yes, the former ones are internal, while the latter ones are external, but why should this be so decisive? As it was pointed out earlier, Searle thinks that only the former ones are significant as far as first-person point of view is concerned. But now, is such a great deal of focus on first-person point of view justifiable? His answer to this question is to be traced back to the manner in which he poses the problem: "The question, in short, is not "Under what conditions does he (Jones) in fact see Sally whether he knows it or not?", but "Under what conditions does he *take himself* to be seeing that Sally is in front of him?" (p.64, italics original) I think this is dubious because it seems that experiences will be left to carry too much cognitive burden. We can imagine a creature who enjoys the same range of qualitative states as humans do, but possesses no cognitive capacities and hence never *takes himself* to be seeing anything. Such possibility certainly weakens the whole intuitive appeal of first-

person point of view. However, this is not deny that experiences are conscious states, that they are, as Joseph Levine (2003) would put it, bits of awareness and that externalists do not have questions to answer themselves. My aim in this discussion of Searle has been to show that externalism is in principle defensible against his objections and that it is possible to move on from (1b), if his body of arguments is taken as a sole basis for supporting (1b). A final remark, a rather interesting one in these circumstances, should be added. One staunch externalist, Fred Dretske (1995), does not feel disturbed by the particularity problem, from which Searle launches his criticism of externalism, at all. His claim is that there is nothing in the experience that tells me whether it is this object, rather than some other object the properties of which are represented in the experience. According to him, experiences have certain informational functions, just as thermometers have the function of informing about temperature. But there is no part of the thermometer that indicates which object it measures the temperature of. Insofar as Dretske's position can be made tenable, it is possible for someone to be an externalist and at the same maintain that experiences do not have particular conditions of satisfaction, in which case an advocate of (1b) would be forced to abandon the particularity problem as a basis for criticizing externalism.

If we accept that the particularity issue does not necessarily undermine externalism either because we believe that Searle's objections are grounded in a false assumption or because the issue may not be an issue after all, and yet want to preserve (1b), then we have at least two more alternatives to consider. The first one is Stephen White's notional content (1994). Its departure from Searle's position lies in contents' not being determined phenomenally. According to White, sensory states do not have the contents they do in virtue of their phenomenal properties. Rather, contents are individuated based on the subject's discriminative skills. Hence this is clearly a relationalist account of content. It is similarity and dissimilarity relations among sensory states that categorise them in terms of content.

White's departure from Searle's views will not be genuine until we've shown how it handles these *prime facie* difficulties: Firstly, it seems that in making similarity and dissimilarity judgements that express the relations in question, a subject relies on the phenomenal properties of his experiences. If it is inevitable that the subject attends to the

phenomenal properties prior to making out similarity and dissimilarity relations among them, why not simply say, like Searle, that content is determined phenomenally? In other words, the mention of such relations would appear superfluous, and thus the relevant discriminations must rely on some other method, even though such method turns out to class the same set of states as similar or dissimilar as the ‘phenomenal’ method would. Secondly, providing the first obstacle is circumvented, imagine that experiences as of something red and as of something green are the entire range of visual states one is capable of being in. A relationalist would be committed to claiming that his green experiences have ‘green’ as their content because they are distinct from red experiences. But this leaves the content of the experiences intolerably underspecified. If another person saw things only as either yellow or red, the answer again would be that his yellow experiences are about something yellow because they are distinct from red experiences. ‘Being distinct from red’ would fail to secure a sufficiently determinate content. Austen Clark’s (1993) is also a relationalist project and the fact that his theory applies solely to the entire spectrum of hues in addition to another two dimensions of brightness and saturation comes as no surprise. Working with the whole spectrum is to eliminate underdetermination as much as possible. What may come as a surprise at a first glance at least, is that a relationalist like White, although not making explicit reference to Clark’s work, gives a clear indication that he wouldn’t regard such project as promising:

Of course, the same object (allegedly) cannot be, and cannot be seen as, completely red and completely green at the same time, colors can be categorized along the dimensions of hue, saturation and brightness, and so forth. But it seems clear that adding these constraints to the intentionalist version of the secondary quality characterization of color properties cannot provide the content that would distinguish red from green. In other words, the holistic characterizations of beliefs and desires seem to involve enough content to define the relevant notions; for colors the opposite is true. (p. 3-4)

Thus White obviously doesn't feel that with his version of relationalism he should be under pressure to make what he calls the 'circles' of colour terms as large as possible, for notional content does not depend on doing so for its tenability. These circles arguably cannot come much larger than three full spectra of hues, levels of brightness and levels of saturation. As it is suggested in the cited text, he nonetheless doubts that they suffice to 'fix' content. Having said this, Clark does manage to overcome the first of the two difficulties above. According to him, colour qualia and, as I will assume, the contents of the corresponding visual state-types are individuated by their locations in such three-dimensional quality space. To oversimplify a little, we might think of this as a structure consisting of slots each occupied by a colour of a specific shade and level of brightness and saturation. Since a state's content is individuated by its spatial relations to all the other states within the structure and not by its phenomenal properties, Clark's account is past the first hurdle.⁶

So how does White's version of relationalism regarding content sidestep the first difficulty, and what makes it independent of the size of the quality space? Central to his proposal is what he calls a file-keeping metaphor. For instance, under normal circumstances, a yellow object emits light of wavelength in band Y, excites receptors of type Y, which results in an image being stored in file Y. The image is not literally coloured, nor do we have anything like files in our heads. We are to think of the image in terms of specific receptive sensitivity, in this case receptors of type Y. Notional content is then assigned on the basis of which type of receptors process the stimulus. So the notional content of the state in the present example is 'yellow' not because it is caused by a physically yellow object, but because the impulse travels via receptors of type Y, or metaphorically again, because it is stored in file Y. This mechanism allegedly underlies our discriminative judgements. To spell this out, another experience is deemed identical by the subject if it also ends up in file Y, that is, if the same receptors (type Y) are stimulated. Now, to contrast Searle's and White's positions, Searle would say that the experience is about something yellow

⁶ I must point out that Clark's quality space is a theory of qualia and not of content. I discuss the problem of content in connection with his theory only for the sake of exposition. White's objection would have some weight only if quality space was adduced to solve the problem of content as well. But Clark's treatment of qualia does not preclude him from proposing a different solution for content.

because it has such and such phenomenal properties; White claims that it is about yellow because it is put away in the same file (file Y) as some other previous experiences and not in files R, B, G, etc. (i.e. it does not excite receptors of type R, B, G, etc.).

Hence the answer to the first part of the above question is that the relations among states described here are not posterior to phenomenal properties in individuating content, they are, rather, what similarity and dissimilarity judgements translate into and, moreover, they are meant to constitute the contents of the states. In the second part of the question we were asking how, on White's relationalism, the content of a state can be independent of the number of other states the state is compared to. In other words, we want to know how notional content accommodates cases where subjects perceive things in just two colours. This does not appear to be a problem since White 'pins' content down on receptive sensitivity. So if someone only has two files, say O and R, the images kept in O, for example, will have a sufficiently determinate content because they are associated with receptors of type O, even though they are only contrasted with a single file (file R) as well as to each other.

In the remainder of this section I shall, once again, adopt the viewpoint of content and qualia externalism, and press a case against 'pinning' content down on receptive sensitivity to account for qualia – a move that commits White to (1b). White, for whatever reasons, follows representationalists in maintaining that qualia are intentional, he is only reluctant to go as far as saying that perceptual content is wide. Indeed, he writes as if the only way for the identity of qualia and intentional contents to emerge unscathed from the Inverted Earth experiments is to abandon the claim that perceptual content is wide. He retells the story of Inverted Earth incorporating notional content into it to show how much better it fares as an account of qualia than wide content. I'll now set out to examine whether the identity can really be preserved only by dropping wide content in favour of notional content. Towards the end of this section I'll also attempt to dispel some other worries voiced by Georges Rey in (1998) who is another author to put narrow representationalism forward as a serious contender. His proposal (set out at more length in his (1997)) only differs from White's in certain details which have little or no bearing on the key points of the

discussion that follows. Where White speaks of file keeping, Rey likes to speak of the characteristic processing of a restricted predicate in the language of thought. But, to repeat, the difference is largely cosmetic, for note that “each type predication is correlated with specific proximal stimulation conditions” (Rey 1997, p.296), which is strikingly in tune with White’s ‘pinning’ notional content down on receptive sensitivity.

Inverted Earth, as considered at face value, is widely held to present insuperable problems for qualia externalism and it appears to be the main motivation behind White’s and Rey’s ‘going narrow’ (Rey in fact lists a number of reasons besides Inverted Earth. I’ll address them below). The thought experiment seems to become notably more benign once we join White and Rey in claiming that qualia are narrow contents. However, I’ll argue that this is a mere semblance and that a thorough evaluation of the experiment reveals that, contrary to what many think of the matter, it is narrow representationalism that it is more likely to topple. It is easy to fall into thinking that retreating to narrow representationalism makes the identity of qualia and contents less vulnerable. Let’s go along with White and suppose that perceptual content is narrow in the sense that it is determined by receptive sensitivity (or, alternatively, by proximal stimuli). The sky on Earth stimulates receptors of type B in Ned’s visual system, therefore the narrow (as well as wide) content of his state is ‘blue’. He is moved to Inverted Earth and has had inverting lenses put in. The Twin Sun (which is blue) stimulates receptors of type Y, meaning that the narrow content of his state is ‘yellow’, therefore, as some like to put it, qualitative differences don’t outstrip differences in narrow contents – there is a qualitative difference, but also a difference in (narrow) contents. As for his wide contents (if, of course, we can show that there are such contents in connection with sensory perception), these are initially false, but we are told that when his use of ‘yellow’ comes to accord with that of the natives, when he starts to use it to mean blue, we have a proof that qualitative differences can outstrip differences in wide contents – his state as he looks at the Twin Sun is qualitatively identical to the one he was in when he was looking at the sky on Earth, yet its wide content has changed.

Since the publishing of Block's (1990), numerous replies have been made to the experiments by representationalists, of which Michael Tye's ones (1995, 1998a) stand out as particularly elaborate. Two alternatives are on offer. The first one draws on his externalist treatment of memory contents (see his (1998b)). Given that it is non-teleological, it grants that Ned's perceptual contents shift alongside the contents of his thoughts. The key idea is that the contents of his past experiences prior to his move shift as well. Consequently, when Ned begins to use 'blue' to mean yellow, and says that the sky is as blue as it was thirty years ago, what he actually expresses is that it is as yellow as it was thirty years ago. The other reply exploits the fact that Ned's receptors have been tampered with, which, it has to be said, arouses rather wittily the odd, gnawing feeling many of us have when first reading Block's paper – that representationalism would actually be worse off if the lenses had no impact on Ned's sensory states. Both replies seem highly plausible and coherent, but I won't go further into them. I'm interested in the teleological response, also placed into the debate by others, one which Tye eventually abandons. Why I want to focus on this particular response is that its alleged failure to handle the problem known as Swampman is, among other things, what drives Rey to embrace narrow representationalism instead. What is more, I believe that it can not only be made effective, but it can also be turned against Rey and narrow representationalism generally. The teleological response begins by disentangling perceptual contents from thought contents in the experiment. It doesn't grant that the former shift alongside the latter. The reasoning is straightforward: sensory states are the result of phylogenetic development, while concepts are acquired in the course of the development of an individual. It would require an evolutionary adjustment in humans as a *species* for the contents to change accordingly. And this would take tens, if not hundreds of thousands of years of further evolution. Rey makes the obvious move - an accidental replica without evolutionary background would have sensory states with indeterminate contents and qualities. Given that 'indeterminate phenomenal properties' is unintelligible, it is concluded that Swampman has no experiences at all. But what argument shows that he would have inner states, let alone inner states that match those of humans? All we are told is that denying Swampman inner life runs counter to certain intuitions. Well,

intuitions are what they are because they are unarticulated, inexplicable hunches. Sometimes intuitions can be supported by compelling arguments, but very often they can't and must be dissolved. Rey, as well as all who take the issue of Swampman seriously, carry the burden of proving that the latter isn't the case. Let's begin with our unarticulated hunch that Swampman would have inner life. We are told that Swampman is a replica of a human being, so the relevant intuition is that he does have experiences and that they match those of a human being. But in what respect is he a replica of a human being? Is he a mere physical duplicate of a human being, or a minimal functional duplicate, or a maximal functional duplicate? It is perhaps in the spirit of the objection that he is physically indistinguishable (a molecule-by molecule duplicate) from a normal human in the first place. Rey's original proposal is that intentional contents and qualia are determined by proximal stimuli, hence he will want to credit Swampman with the same wealth of qualitative states as humans on the basis that he is a minimal functional duplicate of a human. On the teleological response we are considering, Swampman's being a maximal functional duplicate of a human is out of the question. His lack of evolutionary history blocks any description of his phenomenal states which essentially involves a reference to the environment. Should this be seen as the very reason that the teleological version founders? I doubt that the matter can be settled so comfortably, for it could alternatively be seen as evidence that, provided that representationalism is otherwise (i.e. independently of the Swampman issue) the correct account of qualitative states of creatures *with* evolutionary history, there can't be accidental *maximal* functional duplicates of human beings. Thus the teleological version can be vindicated by suggesting the following: "If my theory (i.e. representationalism) is true of humans and sentient creatures *with* evolutionary history (that is, if it's true prior to the Swampman objection being raised), then he is not a *replica* of a human or any other sentient creature with evolutionary history in the most important respect."

Before this is dismissed as too bold and difficult to swallow, I would like to make two points. Firstly, it is part and parcel of the meaning of 'replica' that there is always an intention to copy and, naturally, an original to copy. The Swampman example is devoid of both. All that takes place is a lightning hitting a swamp. There is no

original, no intention to copy. Only if the two elements were present would we be justified in claiming that he is like humans in one particular respect rather than another, that, for instance, he is a physical but not a minimal functional duplicate of a human. In other words, it does not follow from the meaning of the term as it is used in the example, as opposed to ordinary use, that he should resemble humans in one respect and not another - there is no intention to link certain features of real humans with features of our 'replica'. So the dubious intuition under scrutiny really is that Swampman can have 'human' qualia because he can *happen* to resemble humans in whatever accounts for qualitative states in humans. As a consequence, an awful lot will depend on what one already believes to be the correct account of qualia. To repeat, a physicalist will regard Swampman, as far as his qualia are concerned, to be like humans as soon as he is a physical duplicate of a human. A narrow representationalist like Rey or White will regard him (again, as far as his qualia are concerned) to be like a human being if he is at least a minimal functional duplicate of a human being and his physical features will be of secondary importance. Crucially, a representationalist should say, as I suggested, that Swampman is not a replica of any sentient creature in any relevant respect. For if it is true of subjects with evolutionary history that any correct account of their qualia essentially involves a reference to the subject's environment, the possibility of the existence of an artificial replica *with* evolutionary history is necessarily ruled out – any artificial 'replica' *necessarily* lacks evolutionary history and there can be no reference to the 'replica's' environment. If (and only if) otherwise true, the teleological response simply rules out that our protagonist can happen to be a *replica* of a human in the requisite way. It must now be acknowledged that the Swampman objection, by introducing him as a creature that accidentally *replicates* the features that explain qualia in humans, is contaminated with internalistic affinities that make it unacceptably charitable towards positions such as Rey's and White's, while keeping externalism about qualia, quite unfairly, out of contention. It unjustly disqualifies rather than soundly refutes it. And it can't be reutilised by shaking off these affinities, either. What we would be left with would have to be discarded as utterly innocuous.⁷

⁷ Here I part ways substantially with Dretske (1995) who, while holding that qualia are teleo-

Perhaps the case could be recast along these highly hypothetical lines: we imagine that science develops a method that establishes firmly that our Swampman genuinely undergoes experiences identical to ours. Would this reinstate the objection and favour Rey's and White's positions? A persuasive answer to the contrary requires that I connect my last argument with another one: I'll begin with the aid of a little digression and will rejoin the main discussion as soon as its import is clear. It is surely correct to say that the prevalent species taxonomy in biology derives from the theory of phylogenetic development that's along the lines of Darwin's work. For instance, in our ordinary use of the word 'bird', we rely on certain manifest properties of birds, such as the ability to fly, being covered in feathers and so on. At the same time, in certain odd cases, such as when something which has wings, a beak, etc. never flies, it is also true that we ultimately defer to biologists in deciding whether this something really is a bird. The widely accepted classification of species groups together various forms of organisms according to what species they evolved from, what species followed in order of evolution, and what type of environment they have adapted to. It is the structure of their organs, the level of complexity of their physiology, the diversification of the tasks they perform in order to survive and prosper and so forth that are essential to the classifications, and the manifest features such as the ability to fly are significant only as far as they are what these essential features show up as. Hence penguins are birds even if they are far too 'overweight' to ever take off. The use of the word 'bird' is ultimately in accordance with expert taxonomies and thereby the best theory of living organisms we currently have.

Imagine that the biology community have among them a counterpart of the philosopher whose concern about (teleological) representationalism is that there could be swamp replicas. The strayed biologist is contemptuous of the theory of evolution of species, the theory which provides the building blocks of the expert taxonomy, due to the worry that there could be swamp replicas of birds, and indeed of any species biology has thus far recorded. It is clear from what has been said that insofar as it is

representational contents, does not call into question Swampman's being a *replica* and on that basis accepts the consequence that he would have no experiences. My argument has been that the truth of the teleological version of qualia externalism (independently argued) turns the phrase "accidental replica without evolutionary background" into a contradiction.

rooted in the theory of evolution, the current taxonomy makes reference to certain extrinsic facts: organisms are classified as birds because they evolved from reptiles, because such and such species evolved from birds (unless birds are at the end of one of the branches of the evolution ‘tree’) and, finally, because they have adapted to such and such type of environment. These extrinsic facts are an integral part of the meaning of ‘bird’ as introduced by the community of experts we defer to. I think it is reasonable to say that even expressions for certain individual organs such as ‘four-chambered heart’ incorporate such extrinsic facts – having a heart with four chambers marks a certain stage in evolution. Now, our strayed biologist has us imagine that on some distant planet there is a lake that is continuously battered by lightning, as a result of which the lake churns out birds in large numbers (male birds only, I should add, to prevent breeding). Yes, in this scenario all the extrinsic elements are missing, but even if something similar was actually taking place somewhere in the universe, it would be preposterous to see it as evidence that the current taxonomy pertaining to living organisms on Earth (and the theory of evolution upon which it is based) is fundamentally misguided and in need of radical revision. The capacity of our imagination for severing all links to extrinsic factors by confronting us with accidental ‘replicas’ does not render concepts such as ‘bird’ or ‘four-chambered heart’, which incorporate such extrinsic facts by virtue of the nature of the taxonomy they originate in, as empty and detached from reality as ‘phlogiston’ (an imaginary fundamental particle existing alongside electrons, protons and neutrons) or ‘pineal gland’ (wrongly believed by Descartes to be the ‘seat of the soul’, where data from the senses were received) are.

Just as biology identifies certain physiological features that explain differences between, say, reptiles and birds, and in so doing it adduces to the relevant extrinsic facts (one set of these facts are facts about the environment a given species developed in), representationalism identifies environmental stimuli as the factors that explain the differences among qualitative properties of experiences. The concepts ‘reptile’ and ‘bird’, as used by biologists, reflect different stages in the global evolution of species as well as the different environments they evolved in. Analogously, representationalism is a proposal that we view the concepts we use to capture

qualitative differences among our sensory states, such as 'blue', 'green', etc., as reflecting real differences in our environment. That we possess the capacity to envisage physical and minimal functional duplicates in the absence of the relevant extrinsic facts has no implications whatsoever for the claim that the kind of thought contents that derive from perceptual contents, i.e. the contents of thoughts which contain 'blue', 'green', etc., are grounded in such extrinsic facts; the appeal to such capacity is innocuous as a counterargument to representationalism.

One expository advantage of the comparison with biology is that biology is not fraught with the difficulties of having to deal with phenomena which are publicly inaccessible and thus enables us to test the intuitions regarding Swampman and his qualia. Would biologists be reluctant to say that swamp birds are birds? Well, swamp birds would arguably be susceptible to the same diseases, have a proclivity for mating with 'normal' birds, and so forth. Yet, as I have argued, this doesn't imply that the meaning of the term 'bird' as applied to the species on earth should not be grounded in the aforementioned extrinsic facts – adhering to the theory of evolution doesn't compel scientists to be chauvinistic towards swamp birds. It turns out then that representationalism does not lead its adherents to chauvinism with respect to Swampman. They may concede unreservedly that Swampman would be overwhelmed by grief over a loss of his close pal, feel joy or sympathy in other situations, or even enter qualitative states, without recanting the claim that the meanings of 'blue', 'green', etc. as featured in thoughts occurring in subjects with evolutionary history, as well as the intentional and qualitative contents of the sensory states that trigger such thoughts are deeply rooted in the subjects' environment. Furthermore, the human eye is a bodily organ in precisely the same sense as four-chambered heart is. As I have already remarked, four-chambered heart clearly marks a certain stage in evolution and thus integrated into the meaning of 'four-chambered heart' are a number of extrinsic facts, the environmental preconditions for its development in a group of organisms being among them. Since the human eye is just another example of a biological organ, certain form of environmental preconditions (those that led to the development of the human visual system) are present in the meaning of 'human visual system' in a similar fashion. And when this is combined

with the supposition that the human visual system is an instance of an information-processing system, it will only be a short step to realising that these environmental preconditions must also figure in some way in determining the contents of the outputs of the system (i.e. the contents of visual states).

So the upshot appears to be that the representationalist may concede that Swampman can enjoy the inner life of humans without sacrificing any important part of his view. However, this by no means runs counter to the main point of the previous discussion: that he nonetheless can't be a maximal functional duplicate of a human being precisely because, owing to the absence of any evolutionary history, he lacks maximal functional organisation entirely (at least on the teleological version). The semblance of a contradiction is swept clear as soon as we recall that representationalism is a claim that two maximal functional duplicates must phenomenally identical (even when their physical compositions or minimal functional organisations differ). This doesn't commit the representationalist to claiming that, for instance, the experiences of creatures whose maximal functional profile is distinct from that of humans can't feature exactly the same range of qualia as the sensory states of humans. Imagine that evolution has equipped the inhabitants of Inverted Earth with natural inverters to the effect that their sensory states are qualitatively identical to ours. Their states would be qualitatively identical despite their maximal functional profile being different. And, finally, the same holds of creatures with no maximal functional profile whatever; representationalism does not entail that such creatures can't have experiences that are qualitatively identical to ours. The comparison with biology is an alternative route leading to the same conclusion – that Swampman poses no threat to teleological representationalism even if his inner life is as rich as ours.

The argument that teleological representationalism is immune to the possibility of swamp creatures warrants the representationalist's holding on to the initial grievance that the Inverted Earth experiment illegitimately tangles up contents which are phylogenetically fixed with contents which are acquired in the course of individual

development. Inverted Earth does little to motivate a retreat to the positions of narrow representationalism.⁸

On page 30., immediately after outlining White's proposal, I said that in the discussion of Inverted Earth that followed I would adopt the point of view of qualia externalism and that that would enable me to press a case against tying content to proximal stimulation conditions – a move that committed him (as well as Rey, due to his account's bearing the requisite similarities to White's one) to (1b). This is of course in line with my overall strategy in this chapter, which is to wade, step by step, through what lies between (1a) and (3). I also added shortly afterwards that I suspected, and would attempt to show, that the Inverted Earth experiments were more likely to prove damaging for narrow representationalism than for the view which I then set out to defend (that is, representationalism) and which was their original target. I now want to highlight and articulate those consequences of my discussion that substantiate both claims.

It is more than remarkable that Rey and White are silent about the possibility of Block's inverter being implanted behind the retina, or even higher up the stream of visual processing. This subjects them to precisely those problems that they thought fatal for representationalism. The teleological route remains accessible to the externalist regardless of where the inverter is put in. With the inverter in, say, his optic nerve, he can go on insisting:

I am a living creature with an evolutionary history; the environment in which I find myself on Inverted Earth is not my natural habitat, So, on Inverted Earth, optimal conditions do not obtain. The brain state in me that tracks blueness in optimal conditions (and thereby represents blueness) now tracks yellowness. But it does not now *represent* yellowness (Tye, 1995, p.207, italics original).

⁸ I still waive the inclusion in the debate of the two non-teleological replies made by Tye in (1998a) which I think are just as effective against Inverted Earth. More on these replies in the next section.

But (narrow) intentional contents determined by receptive sensitivity and qualia conceived in terms of such contents will in this setting come apart. On Earth, the sky stimulates receptors in band B and a state that is qualitatively blue occurs. The sky on Twin Earth excites receptors in band Y, which results in a state which is qualitatively blue again. Plainly, teleology isn't there to come to the rescue: Clinging on to the narrow intentionalist version, one can not coherently claim that on Twin Earth his qualitatively blue state continues to represent blueness (or, more precisely, that on Twin Earth it misrepresents blueness), despite excitation of receptors in band Y, for our protagonist's receptors do not undergo a change on his arrival on the alien planet. It is only his environment that does so.⁹ So it would be hard to see why a phenomenally blue experience, one that is now caused by stimulation of receptors of type Y should suddenly (mis)represent something as blue, rather than continue to represent something as yellow.¹⁰

That we now find it difficult to anticipate what Rey's and White's reply to this scenario should perhaps be taken to mean that these remarks are not a knock-down argument against narrow intentionalism and that they merely transfer the burden of counter-response back on to their shoulders. But I have already stressed that since they saw no way out for representationalism, we have every reason to be sceptical that any viable effort to salvage their position is forthcoming. At any rate, it seems indisputable now that we are justified in claiming that the Twin Earth examples could prove considerably more damaging for the narrow version of intentional reduction. Rey sums up the position I have been criticising as follows:

I hasten to add that my endorsement of narrow content in cases such as indexicals and qualitative experience does not entail an endorsement of narrow content across the board. The viability of a notion of narrow content

⁹ In the quoted passage, saying that the state "does not now represent yellowness" has the same effect as saying that it now misrepresents blueness, that blueness remains its intentional content, even though the content is false.

¹⁰ A more fundamental reason that Rey and White can't talk of *misrepresenting* and the related notion of falsity is that the type of narrow content they advocate doesn't produce the kind of truth-conditions which would warrant such talk in the present context. They can't distinguish experiences in terms of true/false, at least not in the required way. My current argument can dispense with this point. However, the lack of such kind of truth-conditions will shortly be brought up as a serious drawback (see p. 44).

seems to me to turn entirely on the degree to which there actually exists a stable, internal functional role peculiar to a term (1998, p. 451).

In light of my remarks, it becomes clear that isolating such a stable narrow functional role for every type of visual quality is unlikely to be feasible.

Finally, is there any remaining motivation for tying perceptual content to proximal stimulation conditions even if reducing qualia to such content has probably failed? It appears that neither is there an isolable internal functional role that would at least capture perceptual content. The point that Block sought to bring out from his experiments is that qualia are narrow in the sense that they supervene on molecular constitution. I argued that qualia, when viewed as teleo-representational, and wide contents do not come apart on Inverted Earth, but this defence of qualia externalism did not amount to a refutation of Block's supervenience claim, for, trivially, molecular constitution and qualia do not diverge on the alien planet either. These are the alternatives, indeed, the only alternatives that the experiments license. And note that either alternative (either view of qualia) rules out pinning content down on proximal stimulation (i.e. on internal functional role).

Consider now the first alternative, that is, qualia being supervenient in the way that Block would like them to be. It is *prima facie* plausible to hold that if qualia are thus supervenient, contents can be narrowly functional. But closer scrutiny reveals that, combined with this view of qualia, they can't. Again, on arriving on Twin Earth, our protagonist has an inverter in his optic nerve. The sky excites receptors in band Y (causing a phenomenally blue experience and a thought that, say, the sky is dark blue today). Does he suddenly, for that sole reason, begin to use the word 'blue' to mean yellow? Hardly so. Being unaware of having been transported, the person continues to use it to mean blue until its meaning gets assimilated to that of its native counterpart. Could Rey say, instead, that that the (narrow) intentional content of the person's state remains blue, and that it is only that that content is *false* until his use of the word accords with that of the natives? The answer, again, is negative. Such narrow content is unsuitable to yield the kind of truth conditions which would entitle one to say this. The requisite kind of true/false distinction is inapplicable here.

Consider now the second alternative, one on which qualia are wide. Seeing that if qualia are wide, content can't conceivably be narrow doesn't require an additional argument: if qualia are taken outside, contents will have to go with them – qualia can only be taken outside *as* intentional contents. It turns out now, rather interestingly, that while Block's experiments fail as a rebuttal of qualia externalism, they make at least a strong case for wideness of experiential intentional content. He would certainly not disagree about the latter.

This brings my discussion of Rey and White to an end. I conclude that their appeal to proximal stimulation patterns/internal functional roles holds out little promise whether as an account of qualia or merely as an account of perceptual intentional content. Having dealt with (1a) and (1b), we have now progressed to the final stage of the reversal the completion of which I committed myself to in the opening paragraphs of this chapter. (2), the last obstacle to be overcome, is in fact a slightly more generic version of what Block saw as the only position licensed by the Inverted Earth scenarios. Although not its main focus, the preceding discussion brought to light (3) as a claim that gets an equal licence. It did not, however, go far enough to lead to a refutation of (2). This is the task to be taken up in the next section, at the end of which we should have (3) firmly in place.

1.2 Intentional Content of Experience is Wide, but Qualia are Narrow

To indicate clearly enough how the discussion in this section should resume, I need to be a little more specific than above about what has been achieved so far and at which precise point in the course of our reversal we now find ourselves in view of these achievements. The conceptions that fall under (1b) were irreconcilable with representationalism primarily because of how they approached the issue of experiential intentional contents. Needless to say, their advocates were automatically precluded from agreeing with representationalists on the problem of qualia too, since, to repeat, we can only take qualia outside as intentional contents, discarding any notion of narrow content as we do so. As on all instances of (1b) I have scrutinised content is narrow, reaching (2)

required establishing that nothing of significance (i.e. no arguments those instances rested on) obstructed the route towards conceiving content as externally determined. Searle's self-referential content was first in focus. He was of the view that, by virtue of their phenomenal properties, experiences determine content entirely independently of anything extrinsic. It remains less than clear, somewhat mysterious even, just how they can do so (not least because the phenomenal properties of experience would play a role of unexplained explainers). Towards the end of the discussion it was observed that Searle himself allowed some extrinsic factors (indeed, he could not dispense with them), such as 'the network', among others, to aid phenomenal properties in determining content. This rendered his original claim untenable. Then the focus shifted on internalists (i.e. Rey and White) who, perhaps on the same grounds, recognised that phenomenal properties could not be prior to intentional content in order of determination and instead tied contents to internal functional roles. This strategy, however, was shown to be flawed too.

It appears now that there is no major obstacle to embracing content externalism, which is what (2) partly reflects. What remains to be seen is whether there is anything that threatens to undermine the prospect of embracing qualia externalism. Some of the work has already been done above – narrow representationalism, which embeds sensory qualities in internal functional roles, does not pose such a threat. But rejecting this theory does not suffice to pave the way for qualia externalism: a little later, another internalist position regarding qualia cropped up (purported by Block to be the inevitable consequence of the Inverted Earth story), one which my discussion at the time lacked the resources to refute, namely, that qualia supervene on molecular constitution. Although I defended representationalism against this consequence to the effect that it was no longer seen as inevitable and that representationalism seemed equally legitimate in light of the story, I provided no additional clues as to why representationalism ought to be favoured over it. I shall make this final step in this section.

Peacocke (1983), Shoemaker (1982, 1990) join Block (1990) in accepting the kind of detachment of qualitative contents from intentional contents stated in (2). Although they have been dubbed 'qualia realists' in the literature, suggesting that their opinions emerged from disputes over whether qualia exist or not, I will follow my preferred use of

‘qualia’ and say that they are those who maintain that experiences have qualia as their intrinsic properties.

Before I begin my criticism of Block’s supervenience claim, I want to focus first on what (2) and representationalism agree on, that is, wideness of content. In particular, what I wish to spell out at length is that to hold that qualia are wide amounts to construing them as individuated by so-called long-armed functional roles (or distal or wide functional roles – I will use these terms interchangeably). None of my previous discussions were dedicated to discerning the logical relations that experiential contents enter into through being understood as having a long-armed functional role peculiar to each of their types. Doing so here will, on the one hand, prepare the ground for deciding between representationalism and Block’s supervenience claim (which are, remember, the only two options left at this stage): identifying qualia with such functional roles will seem to be the obvious choice. On the other hand, as in this chapter I am also in the business of characterising externalism about qualia, discerning those logical relations will finally make explicit what qualia are being identified with and thus what consequences we can rightly attribute to the theory. For until now I have mostly been referring to representationalism as a view that external causal factors determine qualia and little has been said to explain that they are only intended to do so through being the input component of long-armed functional roles. They are not claimed to play such a determining role through forming mere input-output relations (that is, in isolation from a subject’s underlying functional make-up), which were central to behaviourism, let alone independently of anything.

1.2.1 Functionalism and Intentional Contents

Functionalism about intentional contents can be formulated in the following way:

(A) All differences in intentional contents are differences in functional roles, and conversely.

The description of the functional role of an experience is fairly complex and involves three types of relations. It relates a state to its perceptual input, to other (relevant) states of the perceiver, and to its behavioural output. Let two experiences be the states of two creatures who belong to the same species and whose visual systems are sensitive to all 'physical' colours and who also have identical functional profiles. Further, the individuals are members of the same linguistic community, which prevents them from expressing different meanings by uttering the same words, as well as from expressing the same meanings by uttering different words. Now, suppose that all we know about their current states is that they have elicited different verbal responses, which we take to be the relevant outputs, and that we have been told neither what the inputs were nor how the states are (functionally/causally) related to other (relevant) states of the individuals. From this we can nonetheless infer that their states are related to different inputs, and, on the other hand, that they also differ in how they (functionally) relate to the other (relevant) states of the individuals. The states will thus have different functional roles in their subjects and, according to (A), different intentional contents. Another alternative to consider is that all we are told about our subjects' states is that they differ in the second type of relations – (functional/causal) relations to the other (relevant) states of the subjects. Again, this much suffices for us to judge that the states have different inputs as well as outputs, thereby securing a difference in functional roles and, if (A) is true, intentional contents.

Note that this is not a defence (or anything of the sort) of the above description of functional role. The three types of relations it involves are simply there and one does not have to be a functionalist to acknowledge their existence. What makes someone a functionalist (or a quasi-functionalist) is a belief that these relations give an experience its qualitative (or intentional – if one endorses at least (A)) identity. My present remarks are only intended to illustrate what trivially follows from the description in the current setting; that a difference in one of the three types of relations results in a difference in overall functional roles of the states being considered. Hence, to sum up, a difference in outputs points to a difference in overall functional roles, and so does a difference in the states' functional/causal relations to other (relevant) states. And, yes, in the example I arranged things in such a way that even a difference in inputs sufficed to immediately

distinguish the states with respect to their functional roles and, more importantly, with respect to their intentional contents, provided that (A) is true.¹¹ In particular, the arrangement that guaranteed this is that the visual systems of the subjects in question belonged to the same species as well as the same linguistic community, and, crucially, had highly idealised discriminatory powers. Therefore, even the slightest difference in the ‘physical’ colours of the samples they are shown produces a difference in the other two types of relations, and thereby their functional roles. The same does not hold for humans - certain subtle variations in the wavelengths of the light that reaches our eyes remain undetected. I will, however, limit our protagonists’ discriminatory powers more dramatically to make obvious the reason why the above description of the functional role of an experience does not imply, except in highly idealised situations such as the one we have here, that distinct (external) inputs necessarily bring about distinct overall functional roles.

Suppose there exists a species of beings who are exactly like us, except that their visual systems are incapable of detecting differences between red things and green things. Furthermore, suppose they see both as red, call them ‘red’, with the rest of their colour vocabulary being exactly like ours. The only word they lack compared with us is ‘green’. Now, two of these creatures are brought together and one of them is shown a red object, while the other is shown a green object. Despite a clear difference in (external) inputs, their states will have identical functional roles – they will be related to other (relevant) states in the same way and trigger the same verbal responses. This is due to the fact that their experiences function in the same way with respect to red things as they do with respect to green things, and are used in thought and action in the same way, regardless of whether they are caused by red things or green things. Such outcome is to be traced to the creatures’ specific underlying functional profiles. It is functional profile that establishes links between inputs and outputs. In this particular case it establishes that a variety of inputs are subsumed under the same response. Hence, knowing that inputs differ is not sufficient to conclude that the states are not functionally identical when this kind of scenario takes place. What is the content of these two experiences? For (A) to preserve its consistency, it must imply that the states have the same intentional contents, otherwise it

¹¹ Throughout the example it has been assumed that the relevant inputs are external too.

would allow for two experiences to be alike in their functional roles and yet have different contents. But what are they? We have seen that it cannot be red for one of the states and green for the other. The most sensible answer would be that the content of each of the states is red or green, or, to put it more emphatically, red-or-green.¹²

Although such sensory organs will probably never evolve in any species, this example is merely to bring out a point which can then be carried over to humans and presumably all other creatures, for there is undoubtedly a threshold, albeit much less conspicuous, to how subtle differences among shades of colours we detect. Now, how does this functional individuation of intentional contents compare to how externalism is typically stated, that is, that intentional contents are individuated by causal factors? This was also, except on several occasions, how I referred to the view before I began to discuss functional roles. The formulation of externalism about experiential intentional contents which does not mention functional roles can be rephrased as follows:

(B) All differences in intentional contents are differences in external/causal factors.

How is the example with creatures who are ‘blind’ to differences between red and green things to be dealt with if (B) is adopted? The word ‘blind’ in the last sentence is of course flanked by quotation marks because it should not be understood literally - as suggesting that there is a real impairment. Every part of our two individuals’ visual systems is intact, which makes them perfectly normal perceivers in the context of their own community. The visual systems reliably perform all the functions they have evolved to perform. These functions include representing both red and green things as being the same colour. Let us consider again the case in which one of them is viewing a red object and the other a green object. The outcome is that they are in identical states. However, some may be tempted to conclude that, on a causal theory, the states will have different intentional contents, given that they are caused by different objects. But this would be a confusion of (B) with this claim:

¹² I do not wish to comment on whether the issue of conjunctivism versus disjunctivism arises here. At any rate, the content of these states can also be expressed as ‘red-and-green’, without this shifting the course of the argument.

(B') All differences in external/causal factors are reflected in intentional contents.

One of the factors that might nurture the temptation to attribute (B') instead of (B) to externalism is that the essential contribution of Putnam's Twin Earth experiments briefly mentioned in section 1.1 consisted in showing that indistinguishability in terms of perceptible properties does not guarantee indistinguishability in terms of contents. Thus the word 'water' as used by Oscar on Earth and Twin Oscar on Twin Earth refers to H₂O and XYZ respectively, despite the fact that their grasp of the concept is rooted in the perceptible properties of the liquids, which are exactly the same. A difference in meanings was explained by an appeal to a difference in external factors. It is certainly true that for our two individuals, the objects are indistinguishable, although one of them is in fact physically red, and the other physically green. The experiences are qualitatively indistinguishable, but are caused by different objects. So can it be objected that the experiences should have different intentional contents if Putnam (and externalism generally) is right? My example, however, is by no means parallel to Putnam's Inverted Earth.¹³ Oscar did not grow up on a planet on which both H₂O and XYZ can be found. Neither did Twin Oscar. Throughout his entire life, Oscar has causally interacted only with one of the substances. And so has Twin Oscar. If, for instance, the rivers on Earth were filled with H₂O and the seas with XYZ, Oscar's concept 'water' would indeed cover both H₂O and XYZ.¹⁴ Only such embellishment of Putnam's thought experiment would make it parallel to my present example. But now the thought expressed by 'The water is incredibly clear' when he stands on the bank of a river has the same content as 'The water is incredibly clear' when he is on the seashore. Thus, there is nothing in Putnam's experiments that tips the scales in favour of (B'). (B') is not the correct statement of externalism.

The task of providing a satisfactory answer to the above question now comes to comparing (A) with (B). Note that, in contrast to (B), (A) does entail that

¹³ Putnam's experiments, of course, address the issue of conceptual content, not experiential content. However, the point about to be stressed does not depend on this difference between the experiments and my present example and can be ignored in these circumstances.

¹⁴ On condition that there would not be a separate concept for XYZ, thus reducing 'water' to mean H₂O.

(A') Differences in functional roles are reflected in intentional contents.

But this feature of (A) has no direct relevance to the kind of comparison in question, mainly because the notion of functional role it includes is further analysed. It is analysed, as we have seen, into three types of relations. We must look at one of the contained relations, namely the one between the experience and its input in order to find the relevant feature. So what interests us is in fact whether (A) entails (B'), that is, whether it entails that all differences in external/causal factors are reflected in intentional contents. This possibility was already ruled out in the example with red/green 'blind' creatures (so this only repeats the point made above), where it was shown that it does not follow from the description of a functional role that

(B'') All differences in external/causal factors are reflected in functional roles.

We only need to substitute 'functional roles' with 'intentional contents' (the substitution is warranted by (A)'s individuation of intentional contents by functional roles) in (B'') to see that (A) does not entail (B').

Regardless of whether we choose (A) or (B) as a formulation of our theory of intentional contents, we will always get the same results after applying them to a variety of concrete scenarios. The only stipulation that needs to be added to (A) is that the kind of functional role it mentions has to be an instance of what I have been calling, together with Block, 'long-armed' functional roles, although this has apparently been assumed all the way through. Long-armed functional roles are roles that have worldly objects as their inputs and outputs, as opposed to short-armed functional roles, the inputs and outputs of which do not reach beyond the boundaries of the body.¹⁵ So if (A) is associated with long-armed functionalism, it amounts to claiming that intentional contents are wide. And this is why Block shares with representationalists his views regarding intentional contents. His functional roles are, again, long-armed functional roles. Hence the agreement between (2) and (3) on wideness of content.

¹⁵ Alternatively, some authors speak of distal and proximal roles, or wide and narrow roles.

Further, we now have a full grasp of why representationalism is a branch of functionalism and also what branch it is. If the representationalist grants that intentional contents are individuated by long-armed functional roles, rather than short-armed ones, and then proceeds to identify qualia with intentional contents so defined, his claim will have exactly the same consequences as (3) ((3) was a claim that intentional contents are wide and qualia are intentional contents). Thus, we can modify (3) as follows,

(3') Intentional contents are wide, for they are individuated by long-armed functional roles, and qualia are intentional contents.

without thereby affecting its tenability. (3') finally articulates what it is exactly that the representationalist reduces qualia to. (3) then should be seen as a compressed version of (3').

Another, and equally acceptable, way of decompressing (3) besides (3') by exposing the functionalist dimension of representationalism is this formulation:

(3'') Maximal functional duplicates must be phenomenally alike.

In view of this alternative statement, which compares entire functional profiles, a number of other features and relations are to be discerned, as so far I have only discussed intentional contents in the context of a single such profile. Some were already adduced to in connection with teleo-representationalism when I, bearing (3'') in mind, raised doubts about the possibility of accidental maximal functional duplicates in section 1.1.2. I said there that representationalism, as stated in (3''), did not imply that subjects with distinct functional organisations could not have qualitatively identical range of experiences, let alone that any of their states could not share their qualitative aspect. Some weighty support for this can be drawn from my considerations in this sub-section.

Let us consider qualia and intentional contents at once. Earlier I envisaged a creature who was exactly like us, except that it perceived both physically red and physically green things as red. While this phenomenal property is what we and the subject in question have in common, a state with such a property plays a different role within that system – a

system with a distinct functional make-up. And since in that system the state plays a role different from that of a phenomenally identical state occurring in us, its content differs too – I characterised it as red-or-green. In sub-section 1.1.2 I also imagined that Inverted Earthlings have natural inverters in their visual organs (ones which they acquired in the course of evolution), in which case their visual states would feature the same range of phenomenal properties as ours, despite their functional make-ups being radically different. As for the contents of their visual states, these would be reversed with respect to those of ours to the effect that, for example, their phenomenally green experiences would be about red things. But I hasten to stress, and this is an important point, that although I called the experience ‘green’, it is not to suggest that there is a misrepresentation - red is what a green experience tracks under the design conditions of their visual system. What follows from this is that even if members of a species whose qualitative life we know nothing about respond discriminatively to the same variety of external stimuli as we do, thereby enabling us to attribute a spectrum of contents which matches ours (i.e. neither would there be any ‘red-or-greens’ etc. to narrow their spectrum in relation to ours, nor any finer discriminative responses to widen it), the representationalist can not extrapolate from human qualia to theirs purely on the basis of having attributed to them an identical range of contents. *Categorise* them in terms of those wide contents is all he can do. He has nothing more informative to say here.

Bringing to light the precise commitments and, no less importantly, exemptions that result distinctively from scrutinising the functionalist dimension of representationalism makes a substantial contribution to my characterisation of the theory, and has at the same time prepared the ground for contesting Block’s supervenience claim. On challenging the claim I shall now concentrate my efforts.

1.2.2 Against Block’s Supervenience Claim

Why should Block’s supervenience be rejected in favour of assimilating qualia to the kind of intentional contents discussed above? To begin, recall his grounds for endorsing the supervenience: he urges that it is possible to implant an inverting device in the optic

nerve, or even further up the stream of visual processing. He then conjectures that a portion of the brain with clearly defined boundaries, one which he calls a 'visual sensorium' could be discovered. Its boundaries would mark the end of the 'transduction' process of an impulse and the beginning of the construction of a visual image. Nevertheless, he adds, this would be of no comfort to the functionalist, since there could be a person with no such clearly defined sensorium, which possibility reinstates his objection. Hence the conclusion that qualia are entirely non-functional, making the case for the supervenience thesis appear compelling. Drawing on his familiar doctrine of continuity of levels of nature in his (1987, chap. 8), Lycan (1996) dismisses the conclusion arguing that the supervenience claim is nonetheless additional. According to the doctrine,

Neither living things nor even computers themselves are split into a purely structural level of biological/physiochemical description and any one "abstract" computational level of machine/psychological description. Rather, they are all hierarchically organized at many levels, each level functional with respect to those beneath it but structural or concrete as it realises those levels above it (p. 118-119).

His thought is that Block's inference to non-functionality of qualia rests on the two-level picture of functional organisation that the quoted passage casts doubt over, one which recognises a single (with however wide or narrow functional roles) software/functional level and a single hardware/structural level, and can only be accepted by those who buy into it. Thus to keep narrowing the functional role that one takes to be peculiar to a certain type of quale in order to escape Block's inverter, only to reach the sensorium with (possibly) vaguely defined boundaries and wrongly find Block's conclusion inevitable is merely to proceed horizontally along the highest levels of functional description. Should there turn out to be such a multitude of computational levels, we could equally well proceed vertically and identify qualia with the functional roles at a lower level. The advantage of this approach in contrast with the two-level picture is that even though Block could successfully attempt to fashion an inversion scenario for each of the higher

levels of description, similar attempts, Lycan points out, would become increasingly contentious as we come near the bottom levels. “For example, we would reach the level of neurochemistry, and the relevant inversion hypothesis would have to be that two subjects could be *neurochemically* identical yet experience different Strange Qualia” (p. 120). Lycan himself does not of course hold that phenomenal properties arise at such a basic level of functional description.

Let me first say that while this line of argument does put considerable strain on Block’s prospects of toppling any version of functionalism with his inversions, it is not fatal, for much to swallow as that particular inversion hypothesis may be, it is not inconceivable. Well, Lycan says that “Block himself, at least, evinces no attraction to any inversion hypothesis of *that* strength” (p. 120). However, I don’t believe that the inversion is out of the question, although it is hardly feasible using an inverting device. It has been established that most narcotics, some of which may affect perception of colours, share a high percentage of their characteristics with neurotransmitters (which certainly do figure in neurochemical functional descriptions) and have an impact on the brain by taking over their *function*. It can be assumed, then, that the inversion in question would be achieved in some such way.

Second, in this sub-section I seek to rebut the claim that qualia supervene on molecular constitution and the argument merely defends functionalism against it. Now, the main reason that I mention Lycan’s doctrine (besides its being profoundly insightful in the present context) and, together with it, the neurochemical level of description is this: There are considerations which show that Block must admit that qualia are functional at least at the level of neurochemical description – and that is enough to make him fall prey to his own objection, given that already at that level an inversion is conceivable.

According to Block’s supervenience thesis, only for molecule-by-molecule duplicates is it necessary that they have the same experiences. As cases where a molecular duplicate of a human differs with respect to his experiences or lacks them altogether are notoriously difficult to bring out, a successful challenge to his position has proved elusive. I will therefore begin with a more modest point: Consider a person who has just been pronounced brain dead but whose heart continues to pump blood around his body. Suppose that his condition was not caused by an extensive mechanical damage to the

brain, but instead by its, say, having been starved of oxygen. The person, of course, can not strictly be a molecule-by-molecule duplicate of a healthy human, as some of the complex compounds in his brain degraded during the time when supply of oxygen was interrupted. Plainly, few would deny that he lacks experiences. Now, is it the mere fact that the person's molecular constitution has been altered that explains the difference between brain dead people and healthy humans as far as their experiences are concerned, or rather that the compounds that are essential in realising mental capacities in humans can no longer perform their *function*? For example, think of neurotransmitters as being among those molecules. While the latter answer is no doubt intuitively more apt, Block would have to insist that the former is the case.

The evidence of multiple-realizability exhibited by the human brain makes the answer I attributed to Block more than just intuitively dubious. Even prior to the findings being available, positions such as Block's were often subjected to charges of chauvinism by philosophers, who protested that on those views some imaginary silicon-based creatures could not become objects of sympathy even if their behaviour strongly suggested the presence of a highly sophisticated inner life. Shewmon's (1999) is a study of hydranencephalics, people whose cortex is severely underdeveloped, or in some cases is missing altogether. He found that, with only their brainstem fully grown, they were capable of perceptual discriminations, a capacity to which only certain regions of the cortex are dedicated in normal children and adults. The chemical description of a hydranencephalic's brainstem when he is in a perceptual state M is hardly identical to that of a healthy person's relevant cortical region when in the very same phenomenal state. So Block now must concede that mental qualities are functional at least at a level where they are already susceptible to inversions, in other words, that the molecular constitution of the brainstem is of secondary importance insofar as those molecules, as it is the case in Shewmon's findings, take over the function of the molecules in the cortex which create perceptual images in normal subjects. Thus the absence of experiences in patients who are in a vegetative state is not down to mere alterations in the chemical constitution of their brains, but rather to the fact that the relevant functional relations no longer obtain. Having said this, it remains to be the case that, even having acknowledged the existence

of a whole hierarchy of functional levels with varying degrees of abstractness, qualia externalism is the only branch of functionalism impervious to inversions.

2. Developing the Challenge

I began Chapter I. with a modest attempt at a diagnosis of our pre-theoretic inclination to regard qualia as intrinsic properties of experiences. The idea was that the entire background of the inclination is given its most faithful expression in a model on which having conscious visual experiences, of both veridical and non-veridical variety, comes down to nothing more than viewing a two-dimensional picture located in the mind. The model would seem to be the default option for anyone uninterested in the science of sensory perception and until some time ago even for many philosophers, since what appears to be going on when - to use that much-worn example once again - I'm having a pink after-image, is that the patch I'm seeing is evidently a part of me, floating there regardless of what's around me or what other states it brings about in me, with pinkness as its phenomenal property, which I simply 'see' that it possesses independently of anything extrinsic. A straightforward extrapolation then fashions the same model for other, less conspicuous kinds of non-veridical perception as well as veridical perceptions. The model squares neatly with the fact that even prior to any schooling in philosophy many of us wondered whether others see common objects in the same colours – if I simply 'see' that pinkness is a property of the patch that I believe to be a part of me, which it 'evidently' possesses independently of anything else, it seems reasonable to think that it may have this property entirely arbitrarily, that is, that others may see it differently under the same circumstances. And it is easy to understand how the preconception stokes the philosophical anxieties (i.e. conceivability of zombies and qualia inversions, direct and infallible access) known as 'the hard problem of qualia'.

This picture of conscious perceptual states finds its most congenial counterpart among philosophical views in a version of the sense-data theory, some details of which were discussed in section 1.1.1. The view marked the starting point in the process of gradually shifting qualia-determining factors from inside to the environment. A version of the sense-datum theory was chosen as the initial stage due to it being a mirror image of qualia externalism, or better perhaps, its reversed form: we have (perceptual) access to the outside world only via the phenomenal properties of a two-dimensional image, i.e. through sensing sense-data; on representationalism, what we have such access to also

determines the qualitative aspect of our sensory states. The two thus lie at the opposite ends of the spectrum. The kind of spectrum that I have in mind, one which I proceeded through in Chapter I. in pursuit of shifting qualia-determining factors outside as noted again just above, is of course only made up of those views of mental qualities that are located in representationalism's immediate vicinity, i.e. those that are interpretable as taking a stand on wideness vs. narrowness of content and whose at least one identifying feature is the way they relate qualia to whatever they believe the correct account of perceptual content to be.

My approach in Chapter I. was largely negative: I simply outlined and rejected several positions in the spectrum until qualia externalism seemed to be the only viable alternative and I may have fallen far short of making a truly positive case for the account I eventually endorsed. But my project in this thesis as set out in the introduction does not demand that I do so. This should become much clearer when the main line of argument starts to develop later in this chapter. Suffice it to say now that all that was required there was, firstly, to show that a move from sense-data and then along our entire spectrum to representationalism can succeed, and, secondly, to extract in the process a number of essential characteristics of representationalism to be adduced to in the remainder of my work.

Mentioning at the outset the pre-theoretic pictorial model of perception and its philosophical counterpart – the sense-data theory – will help introduce the subject matter of this chapter. Among the positions considered in Chapter I. the sense-data theory is the odd one out in a sense that is crucial for my purposes here. It was presented as falling under (1a), on which sensory experiences make us aware of internal occurrences and so their intentional contents as well as qualitative properties are narrow. The feature that sets the view apart from the other ones considered in chapter I. is that the kind of intentionality involved in (1a) is in fact higher-order intentionality, for those who once held a version of the theory would be reluctant to say that sense-data are intentional. The source of higher-order intentionality here is the fact that sense-data themselves are sensed – they do not constitute perceptual awareness, rather, it is the 'inner eye' that does so. By contrast, the other competitors of (3) that I dealt with after dissolving mental objects of perception with a Lycan-style counterfactual analysis only involved intentionality of first-

order variety. Well, some may protest that sense-data do not generate genuine higher-order contents just by being targets of sensing – they are to be conceived of as (mental) objects rather than first-order *states* and thus it would appear less than reasonable to speak of higher-order intentionality in absence of first-order states; the theory, as the objection might continue, merely transfers objects of sensing inside the subject (or more precisely, to the mind). Similarly, the fact that my aching stomach is inside me does not make the relevant pain-state a higher-order one. Though I grant that there are less contentious instances of higher-order states (which is another testimony to how difficult it is to come to grips with sense-data theories), we do have an overriding principle to invoke which rules that the state of sensing a sense-datum falls into that category after all. Note that sense-data are purported to carry qualitative properties and are therefore, unlike my aching stomach or common worldly objects, items belonging to one's psychological life. Hence higher-order states ought to be defined as what facilitates the mind's awareness of what it comprises.

Since sense-data are by no means first-order *states* and yet generate suitable higher-order contents, such arrangement would perhaps be most appropriately labelled as a 'proto-higher-order account' of conscious experience. In other words, it's an early version of - and this is what I aim to highlight in relation to the central focus of this chapter - detaching the state which constitutes perceptual awareness on the one hand, from the qualities-bearing component on the other. The objective in this chapter is to argue that the detachment is a substantial leap forward and that any viable conception of the architecture of phenomenal consciousness must incorporate higher-order awareness into it. However, once we have this point firmly in place, it will transpire that the only plausible candidates for providing vehicles for higher-order awareness (those directed at lower-order qualitative states in particular) are natural-language sentences, the semantic properties of which derive from the role they play in inferences. Consequently, the kind of semantics we shall adopt for what (i.e. natural-language sentences) lends states of awareness their shape, structure and, most importantly, their identity will inevitably produce a view on which the only genuine form of awareness we know of and recognise is one which amounts to a move in a 'logical space of reasons'. Of course, at this stage in the argument, Sellars will already be engaged in the debate. In the latter sections of the

chapter I will explain how this approach to higher-order awareness, even when paired with externalism about qualia to provide a complete architecture of phenomenal consciousness, collapses into something which very closely resembles the sense-data theory of sensory perception. Rather dauntingly then, the present paragraphs are not the last time I've mentioned sense-data; and when they are reinstated in the way I've just indicated, it will seem utterly futile to seek to dissolve them with Lycan's counterfactual analysis again and proceed towards qualia externalism, for we will risk getting entrapped in the very same circle.

2.1 Worldly vs. Experiential Subjectivity

Externalism about qualia - the position we arrived at in the previous chapter - was stated as follows:

(3) Experiential intentional contents are wide and qualia are intentional contents.

The standard definition of qualia – or at least that endorsed by “qualia realists” - is that they are intrinsic properties of experiences in virtue of which experiences are conscious. And an experience counts as conscious only if there is something it is like to have it. I said already at the beginning of chapter I. that my use of the term throughout the chapter would not entail that they were non-relational and that my discussion would be that of their status (intrinsic vs. relational) rather than their existence. However, for the sake of the overall argument in chapter I. as well as the present one, the second part of the definition was retained in that qualia were understood as what accounts for “what-it’s-likeness” of experience’ (see page 1.). On such understanding of ‘qualia’ in (3), the claim can only be attributed to first- as opposed to higher-order representationalism. For convenience, following other works on the subject, the acronyms FOR (first-order representationalism) and HOR (higher-order representationalism) will also be adopted.

If chapter I. was a first point of departure from the “qualia realist’s” construal of qualia in that the most plausible view out of those considered there (i.e.

representationalism) diffused intrinsic qualia in one's maximal functional network, then in this chapter a further deviation from it will be encouraged, one which leads to denying that a state's possessing qualia is a sufficient condition of phenomenal consciousness, thereby paving a way towards a HOR theory. But I must stress that although having ditched the belief that qualia explain "what-it's-likeness" of experiences does amount to rejecting first-order representationalism, many fundamentals of it will be built into the resulting higher-order view: once we have stripped qualia of phenomenal consciousness as HOR theorists recommend, the residue (I will also call it 'residual qualia'), taken now only as a necessary condition, will still be identified with wide contents. Having said this, I by no means wish to recant the claim I made at the end of chapter I: that out of those positions to which the way they relate qualia to first-order contents (with the exception of sense-data theories, which, due to their anomalous nature, presuppose higher-order contents) is somehow essential or are at least so interpretable, first-order representationalism is by far the most acceptable one.

Carruthers (1996, 2000), Lycan (1996), Armstrong (1984), Rosenthal (1986, 2005), and Dennett (1991) are the most prominent authors who hold or in some way sympathise with higher-order approaches to phenomenal consciousness. Only the first two can rightly be regarded to be intentionalists about qualia (understood here as "stripped" of consciousness) carried by first-order states. Carruthers is in fact an internalist about these, appealing to the type of functional roles (of first-order states) that Rey and White (see section 1.1.2) thought generated phenomenally conscious experiences. Incidentally, the ground that Rosenthal occupies in his (2005) at the first-order level bears a striking resemblance to Austen Clark's (1993) quality space touched on briefly in 1.1.2 too, although he prefers to call it a 'homomorphism theory'.

Since it is intentionalism regarding consciousness that I'm concerned with in this work, we will set aside the varying details of these authors' approaches to first-order representations, and whatever turns out to be the most suitable candidates for serving as vehicles for higher-order states will be conjoined solely with externalism about (residual) qualia so as to provide a complete picture of phenomenal consciousness. And, to reiterate, I already went to some lengths to argue for externalism in the previous chapter, even though there qualia were not bereft of the conscious component. But again,

essentially the same arguments could be deployed to support externalism about residual qualia.

Thus in most generic terms, higher-order approaches to consciousness are ones on which a conscious experience comes into existence over two isolable but more or less coinciding stages: a qualitative one and a conscious one. Now, showing the following claims to be true would facilitate mounting a powerful case against first-order representationalism and for higher-order representationalism at the same time: (i) that there are qualitative states that occur in absence of consciousness, (ii) that there are states that are genuine instances of perceiving, to which the FOR theorist would have no choice but to attribute wide *perceptual* contents or face charges of inconsistency, but which nonetheless lack phenomenality altogether, (iii) that HOR has the resources to explain the difference between states such as those described in (i) and (ii) on the one hand, and conscious experiences on the other, and finally (iv), that only a HOR theory and not first-order representationalism is in a position to explain why my experience as of something yellow should feel like this rather than like something else. That (iii) and (iv) are included is to pre-empt objections that, even if proponents of HOR do manage to substantiate (i), adding HOTs (higher-order thoughts) would shed light on just one kind of consciousness, that is, what Block (1995) calls access consciousness, while it leaves the problem of phenomenal feels untouched. Furthermore, even (i) seems highly unintuitive at the first glance, for many would take the view that what is distinctive of qualitative states is that access consciousness is inherent in them – it is not conferred upon them from outside, whether from upper levels or elsewhere. Positing HOTs, they would urge, is a plain redundancy, a duplication of what is already installed in experience anyway. Thus Levine writes:

The inadequacy of both HO and the access/phenomenal distinction manifests the paradoxical duality of qualitative experiences: there is an awareness relation, which ought to entail that there are two states serving as the relevant relata, yet experience doesn't seem to admit of this sort of bifurcation. Let's call this the problem of "duality." That qualia have this dual nature, and that certain conscious

thoughts are phenomenally constituted, are clearly intimately connected. (2001, p. 168)

Carruthers clearly acknowledges that addressing these worries satisfactorily must be among the *desiderata* for HOR:

For why should an analog, but non-conscious, perceptual representation suddenly acquire the subjectivity distinctive of phenomenal consciousness merely because it causes a higher-order belief about itself? How can the mere fact that I have non-inferential knowledge of the occurrence of a certain experience make it the case that there is suddenly something which it is like to undergo that experience? And how can the phenomenally conscious differences between distinct phenomenally conscious states be explained, on this account? For it looks as if those differences can only be differences between the contents of the first-order states targeted by HOTS in each case. The difference between having a HOT that I am undergoing a state with the analog content red_a , on the one hand, and having a HOT that I am undergoing a state with the content $green_a$, on the other, can only reside in the differences between the first-order contents red_a and $green_a$, but these differences are already there in the non-conscious perceptual states which become targeted! (2000, p. 239)

Carruthers' body of arguments covers all of the above four points. As for (ii), he is the only higher-order theorist who invokes some compelling evidence for perceiving without phenomenality reported in cognitive science literature to argue for a higher-order account. Lycan (1996) and Rosenthal (2005) would have something to say about (i), (iii) and (iv), but these alone would surely suffice for FOR to be superseded by HOR.

In the remainder of this section I shall look at what justification for (i) – (iv) is provided by advocates of HOR and in so doing I will seek to elucidate how they jointly lend themselves to Carruthers' distinction between worldly and experiential subjectivity, and the more or less parallel qualia/"what-it's-likeness" and thin/thick phenomenality

distinctions drawn by Lycan and Rosenthal respectively. These distinctions will then give the idea that phenomenal consciousness emerges at the second-order level a very solid footing.

Rosenthal's, Lycan's and Carruthers' arguments for (i) are in large part a reflection on what Armstrong (1981) termed 'minimal consciousness', describing experiences of an absent-minded driver. This is a common phenomenon, one that is particularly likely to occur when we go about the most tedious of our daily routines, when our thoughts drift away from what we are currently doing. So Armstrong's long-distance truck driver may happen to be day-dreaming while he continues to drive safely, negotiates a roundabout, stops at traffic lights and follows road signs. Rosenthal mentions prolonged headaches to bring out the same point. A headache can last for hours, but we repeatedly forget about it or something sufficiently entertaining takes our mind off the discomfort. Despite feeling a momentary relief on such occasions, we would not say that we had a "sequence of discontinuous, brief headaches" (p.39). These are just some of the familiar cases pointing towards a dissociation of qualia from conscious awareness. One might still ask: "If you are unaware of a quale, how do you know it's there? Well, consider the long-distance driver whose thoughts have drifted away from driving. It is certainly correct to say that, as long as his eyes stay open, his visual system continues to process stimuli and create visual images, and that these stand readily available regardless of whether his thoughts wander around. Were this not so, one would be obliged to claim that when the driver chooses to concentrate on driving again, his attention mechanism somehow issues an instruction to the visual system to resume construction of images. But during the time when he drove absent-mindedly, he acted on what his visual images presented him with anyway.

Although explaining the significance of these examples for HOR will be postponed until we get to (iii) and (iv), it is immediately obvious that they, and the fact that we form thoughts about and *act* upon the contents of non-conscious perceptions in particular, will prove awkward for the FOR theorist. Recall from section 1.2.1, where I examined in some detail the consequences of functionalising contents, that a state's intentional link with external objects (i.e. its wide content) is set up by the unique long-armed functional role which that state-type occupies, for if an account of wide content failed to involve

functionality and instead attributed contents solely on the basis of (external) inputs and responses, the result would be a simple input-output behaviourism. For that reason, (3) was said to be equivalent to the following:

(3') Intentional contents are wide, for they are individuated by long-armed functional roles, and qualia are intentional contents.

Now, tokens of the same experience-type are invariably accompanied by certain other mental states as well as actions, beliefs and verbal reports, many of which are classifiable as either first- or second-order. Functionalists are typically selective about which of these causal links should make a contribution to defining a state's functional identity, as some are deemed marginal or entirely irrelevant as far as its role in the system is concerned.¹⁶ Which get selected of course very often varies wildly from author to author, but of those states, actions, etc. which are classifiable as either first- or higher-order, the first-order representationalist, whose claim (3') is, will only be keen to include certain members of the former category, otherwise he would adulterate his account with elements of HOR. Therefore the P in Tye's PANICs stands for 'poised to have an impact on *first-order* beliefs and actions'. The problem then is that quite irrespective of which first-order states, actions, etc. are granted the privilege of forming constitutive functional relations to, say, my conscious experience as of something pink, it is hard to see how at least some of those relations could fail to obtain when an experience of the same type occurs non-consciously.

Imagine that while the truck driver drives absent-mindedly, there is a person in a pink tracksuit crossing the road. With his thoughts far away, he brakes to let him cross. It makes perfectly good sense to say that this (non-conscious) first-order action resulted from a (non-conscious) first-order belief that there is a person in pink crossing the road, and yet neither of them guarantees that the visual state that had caused them was conscious. Even an appeal to functional relations to first-order verbal reports would not

¹⁶ For instance, a processing unit in a computer is usually connected to a indicator light which tells the user when it is in operation, but if the computer had been built without the light, that would not have altered its role in the system or made it any less clear. It is relations to certain other components that define its role as a processor and hence the functional term 'processing unit'.

do, however strongly we are inclined to think that we can only report those experiential contents that we consciously attend to. Suppose that the driver, while, for example, overcome by sorrow at the thought of some tragic recent event, non-consciously misjudges his distance from the pedestrian in proportion to his speed, so when the truck suddenly screeches to a halt, he shouts: "Thank heavens I noticed he was too close and started braking harder!" Though he would surely have 'come round' by this point, the utterance may well originate in a visual state which occurred prior to his coming round, at the time merely causing him to think non-consciously that the man was too close and that he had to brake harder. For when later asked about when exactly he noticed that the man was too close, which he must have done judging by, say, the skid marks, he might answer that he doesn't remember because he was still overwhelmed by memories of the tragic event and that it was the spontaneous verbal reaction that really drew his conscious attention back to the dangerousness of the situation.

True, few would be prepared to accept an account which makes the presence of phenomenal consciousness contingent upon the capacity to produce verbal reports, but it is considered here to help emphasise the fact that the search for such functional links to first-order states that would clearly differentiate between conscious and non-conscious instances of the same experience-type is futile. All conditions for the presence of phenomenal consciousness that the representationalist can legitimately put forward within the confines of a first-order theory are equally well satisfied by the kind of non-conscious states described above.

It is easy to unwittingly slip into thinking that here I simply overlook the fact that it is the absence of first- and not higher-order attention that explains why the driver's experiences are non-conscious. His attention - and this is surely correct - shifts from the *contents* of his experiences rather than the experiences themselves to, say, his empty stomach and thoughts about food. Does this mean that the situation can be dealt with in purely first-order terms by claiming that only those perceptual states are conscious whose (wide) contents become targets of first-order attention? To ask this is to already lose sight of what has just been said in connection with (3'). According to (3'), an experience's wide content (attributed on the basis of its long-armed functional role) accounts for what it is like to have it. Bearing this in mind, if the first-order representationalist believes that

the answer to the question is 'yes' and takes the view that there is nothing it is like to have an experience when it is not accompanied by first-order attention, then the consequence to which he commits himself is that the driver's experiences somehow cease to be *contentful* states the moment he stops concentrating on his driving. If they didn't, wide contents and what-it's-likeness would come apart, thereby undermining the identity of contents and qualia asserted in (3'). The only way to avoid the consequence would be to make attribution of content conditional on the presence of first-order attention by integrating it into the functional roles peculiar to each perceptual state. But this is a failure to appreciate the significance of the notion of contentfulness in theorising about mind - a failure to appreciate why we attribute content at all. The notion of contentfulness is intimately connected with behavioural discrimination. To say that a state is contentful is in effect to credit its subject with the ability to show discerning behaviour with respect to the environment. It is difficult to see how attention could figure in this conceptual connection and be decisive in drawing a divide between contentful and contentless states.

Followers of Block's typology of consciousness would take a different tack and, as it was previously anticipated, protest on behalf of FOR that states with unheeded contents merely lack access consciousness not phenomenal consciousness. On such (terminological) rearrangement, the fact that much of the behavioural pattern unique to a certain state is preserved in absence of attention would prove favourable for FOR. The move, however, is forestalled by (ii).

(ii) was a claim that there are states that are genuine instances of perceiving, to which the FOR theorist would have no choice but to attribute wide *perceptual* contents or face charges of inconsistency, but which nonetheless lack phenomenality altogether. The claim, provided that its truth is beyond doubt, will bring out another sense in which an experience can be non-conscious. The perceptions of the absent-minded driver were non-conscious due to (first-order) attention-shifts, but they possessed phenomenality - their contents continued to be accessed (intended here more broadly to include access by non-conscious thoughts; not to imply the sense of 'access' intended by Block) as contents of qualitative states. Disputing this would mean that the visual system ceases construction of images whenever attention shifts elsewhere, or so I briefly argued. (ii), by contrast, is about perceptions whose contents are not accessed as contents of qualitative states,

whether the kind of access in question is achieved through attention or when in absent-minded mode.

I have already remarked that it is Carruthers (1996, 2000) who, from the perspective of HOR, has recast a set of findings reported in cognitive science literature (particularly (Milner and Goodale, 1995), (Marcel, 1998) and (Weiskrantz, 1986)) as an argument for (ii). In the next few paragraphs I shall follow how the evidence is transcribed in his work into the terminological framework that is home to the kind of philosophical issues of consciousness which are currently under scrutiny. To begin, he explains that stimuli picked up by the retina are handled along not one but two distinct routes. However, contrary to what was previously assumed, the information travelling through the two transduction channels does not end up being integrated to form a single visual representation. Rather, the streams terminate in spatially different, and, more importantly, functionally dissociable regions of the brain. We can dispense here with the neurophysiological details about the two channels of processing and the precise locations of the corresponding terminal regions and concentrate instead on their functional description in Carruthers' text, as this will help make intelligible the idea of perceiving without phenomenality. Distinctive of one of the channels is the fact that its stimuli flow towards the regions believed to be dedicated to generating visual images (in the temporal cortex), where those states are subsequently available for conceptualisation and reasoning. The areas associated with language are also thought to be linked to these regions on the functional map of the brain. On the other hand, the second route leads to the evolutionarily inferior parts of the brain designated on the map to control instantaneous motor responses (in the parietal cortex), such as grasping or catching a moving object.

Based on the copious dissociation data he rehearses, Carruthers feels entitled to speak of the processing paths and their terminal regions as discrete visual systems. Predictably enough, the end products of the former are what philosophers of mind would label "qualitative", while the system which controls execution of swift motor reactions involves no qualities at all. The close connection between the 'qualitative' system and the concept-wielding faculty and language on the functional map of the brain is interpreted as suggesting that qualitative states are mainly required for deployment of concepts in

identification and recognition of objects, which on many occasions involves lengthy examination and assessment of the perceived scene. The motor system comes into play when, for example, the object that has just been identified is in motion, and the need arises for whatever reason to alter with great speed and accuracy the position of one's body or limb in relation to the object. So when trying to stop a mosquito from disrupting my sleep, the 'qualitative' system facilitates my identification of that thing as the insect that is being the nuisance, but then, once it's been identified, it is the motor-control system that assists my attempts to follow its frantic movements with my hand as I prepare to swat it. Carruthers mentions taking part in fist-fighting or sprinting over uneven terrain as cases when the latter is engaged.

Besides the hypothesis that only object-recognition and not instantaneous and accurate execution of motor responses should be aided by qualitative states, there are reasons for the motor system's lack of phenomenology which have more directly to do with the purpose it has been designed for. It seems natural, or so we are told, to think that one's ability to adjust effectively the position of one's limb in relation to a moving object is enhanced if his motor system by-passes, so to speak, construction of images and its functional accompaniment – conceptualisation and reasoning. Of course, delivering sensory images would in itself cause at most a negligible delay even when the swiftness of accurate motor reactions is at stake. But I think the following analogy will bring out nicely why qualitative states should nonetheless be an impediment: At the most basic level, operations in computers are carried out in the so-called machine code. The units or symbols over which it computes at that level are sequences in binary code. The computer also has an interface, whereby certain information is available in the form of a graphic layout on the screen to be utilised by the user. But regardless of how the information is graphically presented to the user, the computer stores and handles it as sequences in binary code. Corresponding to such information in the visual system would be stimulation patterns moving up the 'qualitative' route while carrying information about the environment.

As it could be guessed, I want to suggest that qualitative visual states (i.e. visual images) are in their own characteristic way the brain's interface with the mind (this is of course not to indicate an ontological split between the brain and the mind). The first point

to be made is that visual images cannot directly cause action, prior to some aspects of their contents being conceptualised, otherwise it would be utterly mysterious why some particular perceptual contents result in this and not some other course of action being taken (or alternatively, why only some and not other aspects of an experience's contents are acted on at a particular time *t*). Plainly, I can attempt to swat something only if I have identified what I see as the insect that is causing my annoyance. But if I instead identify what I see there as a speck of dust that's gone airborne, a different course of action will be taken – I will simply look elsewhere for the mosquito I am after. So qualitative states lead to action only insofar as they are transformed into perceptual beliefs, thoughts, etc. In other words, they are only available to be utilised in or via conceptualisation and belief-formation. I mean 'belief' in a broader sense here to include the perceptual beliefs of the absent-minded driver or even beliefs made up of more primitive concepts attributable to animals. When the absent-minded driver steers around a hole in the tarmac, he is credited with the belief that there is a hole in the tarmac. To be more precise, and this is another point to highlight, he is in fact credited with a whole set of beliefs at once, for example, that driving over holes wears the suspension, that he will spill his drink if he doesn't avoid it and so on, as well as, needless to say, the possession of a wealth of interdependent concepts. The same arguably holds for all cases of exploiting the contents of qualitative states – in order to be of any use at all to 'the user', they must bring into play not just one belief, but an entire network of inferentially connected beliefs (and concepts). Similarly, no further processing of the information shown up in the computer interface will take place until the user decides what to do with it, and what decision is made will very much depend on his current needs and beliefs.

In light of these points, it is understandable that the visual system should delegate the task of controlling instantaneous or instinctive motor reactions to a subsystem which bypasses construction of images. In short, qualitative states cannot result in action without a body of concepts and beliefs having been applied, and with them they are much to go through before their contents are acted upon. Perhaps the absent-minded driver's situation is fairly straightforward to deal with, and, in addition, steering around the hole by continuously conceptualising the changing scene as he approaches it at a normal speed would not be a hindrance.

But consider, as Carruthers does, catching a ball, especially when it is thrown with power and takes a deflection, or an awkward bounce as it spins. We often speak of instinctive reactions, and rightly so, given that every time a ball is thrown in this way, it comes at a different speed, has a different trajectory, spins differently, etc. It is questionable whether we can even conceptualise and put together in the form of beliefs the countless variables and their values according to which we adjust the position of our hands, and which values, let alone their combinations, are never repeated with a new such throw - not to mention the cognitive burden on us if we could and did; and yet in most cases the adjustment is well-timed and made with remarkable precision. We would feel strongly here that the only explanation for such a high rate of success is that the visual system must, in the proper sense of the word, take account of those innumerable factors in guiding the movements of our limbs - it is only unlikely that the subsystem which allegedly guides those movements should make them and their values available, by displaying them in an 'interface', to 'the user', that is, to the conceptualising a deliberating self.

These remarks are primarily intended to outline the role of non-qualitative perception within a sentient creature provided that there does turn out to be such mode of perceiving. It can be taken as a sketch of an argument for its existence only if it is underpinned by weighty experimental evidence. Without it, many would consider it vague and largely anecdotal. Carruthers draws on a number of studies in which one system is dissociated from the other in various impairments, and even experiments with healthy people which point towards this qualitative/non-qualitative division in the visual system. I will focus on just one clear-cut case where the ability to act on perceptual contents is preserved while phenomenality is partially missing – a now familiar disorder known as blindsight.

To begin, note that there are other examples of sensory 'representations' without phenomenality such as those involved in early visual processing. Neander (1998, p. 412) takes HOR theories to be partly an answer to the question as to why those representations aren't conscious. It would be a mistake, however, to think that this question alone calls for a distinctively higher-order answer. Representations built in the early stages of the processing of a stimulus fail to satisfy at least the P in Tye's set of requirements abbreviated to PANIC. They are immediately followed by further processing of stimuli,

not by beliefs and actions. The stipulation that qualitative states be poised is to mark the point where the non-conceptual and the conceptual meet – the point where representations acquire phenomenality. This is akin to the idea laid out above that qualitative representations are in their own characteristic way the brain's interface with the conceptualising self.

As in the case of representations constructed in the early stages of processing, Tye's response to the facts of blindsight is that the states they undergo in their blind region aren't suitably poised; they aren't apt to produce beliefs in their subjects, since, on being prompted, blindsighters merely guess what's before them and are taken aback when they learn about the accuracy they achieve. Although he is right that they do not form beliefs with respect to those states, Tye's notion of poisedness (as well as the rest of the requirements), I will argue, fails to pinpoint the difference between blindsighters and normally sighted people. It fails, that is, unless he wants to be understood as saying that beliefs, or, rather, dispositions to produce beliefs somehow endow perceptual representations with qualitative status. A further elucidation of his notion of poisedness will reveal why this is so.

Consider first the following passage:

... representations are built up of distal features of the surfaces of external objects in mechanical fashion by computational processes. The initial, or input, representations for the visual module track light intensity and wavelength, assuming nothing is malfunctioning. The output representations track features of distal stimuli. Thereby, it seems plausible to suppose, they represent those features. It is here that things initially acquire their looks, here that basic perceptual experiences are found. Likewise, for the other senses. (1996, p. 293)

The passage summarises the computational processes which take place in the visual system. The computations are performed automatically, independently of the upper levels of cognition in the subject. The early symbolic representations that are constructed carry information about *proximal* stimuli and are converted over several intermediate

computational steps the character of which need not concern us here into symbols carrying information about *distal* stimuli. Sets of symbols become phenomenal states the moment they come to represent distal features, the moment they acquire wide contents. Note that there is no mention of beliefs or actions yet.

There is, however, a second line of thought and it is here that the functional link to beliefs and actions that he calls 'poisedness' comes into play. Later in the same article Tye speculates about which simple creatures have states with PANIC, in other words, which simple creatures can be plausibly thought to entertain qualitative states on the PANIC theory. His conjecture is that the presence of suitably poised states in creatures lower down the evolution tree is invariably manifested in behaviour with a certain degree of flexibility. What underlies flexible behaviour is the ability to access, by means of forming beliefs, the contents of one's perceptions and discard certain other beliefs in light of those contents. Thus Tye has it, for instance, that caterpillars' states lack qualitative aspects. When foraging, they move, without exception, towards the brightest light source. If a sufficiently bright artificial light with nothing edible around it is placed nearby, they will be attracted to it and stay there until they starve to death. There is no learning and therefore no behavioural flexibility to warrant attribution of beliefs (e.g. beliefs that there is a light source with such and such a degree of brightness). To put it in the spirit of the quoted passage and show how this second line of thought connects with it, since all their arguably primitive visual system registers is light intensity, i.e. features of proximal stimuli, they cannot be said to act upon or have beliefs about wide perceptual contents; the computations of what is initially proximal stimuli do not reach the stage where representations of distal stimuli are generated. This must not to be confused with the case of a person whose vision is so severely impaired that he can only make out crude differences in brightness between objects, for even though his discriminatory powers would not be far superior to those of, say, caterpillars, he continues to have widely contentful representations, albeit substantially impoverished. The motion of the caterpillar's body, by contrast, is the result of a brute-causal connection with the features of a proximal stimulus.

Compare this with what Tye says about fish. Some predatory fish can, for example, learn to avoid artificially coloured fish that have been injected with a chemical to make

them unpalatable despite otherwise being included in their diet. Having mentioned several other experiments, he concludes that their behaviour exhibits the kind of creativity or flexibility which can only be explained by an ability to form novel concepts and beliefs (e.g. such and such is no longer edible), and that the formation of those must be facilitated by their genuinely examining or assessing a scene as it is laid out before them – as opposed to displaying mere brute-causal input-output transitions.

So here is why, at least on my reading of Tye, the term ‘poised’ in his theorising about phenomenality is intended to mean ‘poised to make a difference in *beliefs* (and actions)’ rather than just ‘poised to make a difference in actions’: the former expresses the thought that when a representation of distal stimuli occurs, the transition from the (external) input to behavioural output is never brute-causal, but is instead mediated by beliefs. And although such representations do not always cause beliefs with respect to their contents, the fact that they are so poised is purported to capture something which is essential to qualitative states, something which the latter doesn’t, for it would apply equally well to the visual states of caterpillars or to yet simpler organisms that react to light in even more primitive ways.

I very much agree that ‘poised to make an impact on beliefs’ thus understood succeeds in capturing the difference between us and far simpler organisms whose visual systems are not sophisticated enough to generate spatial representations. Nevertheless, I shall argue, it falls short of identifying what is essential to qualitative states precisely because it does not differentiate between us and blindsighters. Carruthers calls attention to experiments reported in Marcel (1998) where blindsight subjects attempt to reach out to grasp objects with somewhere between 95 and 98 per cent of normal precision.

Think how sophisticated the perceptual processing must be in these cases. When a blindsight patient reaches out for a cup placed at a certain distance on the desk beside him, he has to be able to estimate the size, shape and orientation of the object, as well as its distance from him. But all this is done, remember, while the patient thinks he is guessing randomly. (2000, p. 156)

Let me first stress that my interpretation of Tye's notion of poisedness differs from Carruthers' in that he is prepared to grant that the states blindsighters undergo in their blind region aren't poised (to make a difference in beliefs). But, according to him, this just shows that the functional connection with beliefs is superfluous. Given that, as it is clear from Marcel's findings, blindsighters' perceptual discriminations are almost as fine-grained as those made by healthy people, he wonders how the absence of beliefs could account for the lack of qualitative states in the former, or, how their presence in us could turn those states with similarly rich contents into qualitative ones.

Although, again, Tye's response to the facts of blindsight is that they are not poised since there are no genuine perceptual beliefs, I doubt that he wishes to be understood as maintaining that beliefs impart phenomenality to otherwise non-qualitative sensory states. Rather, it follows from the above elucidation of 'poisedness' that its role in Tye's reasoning is to mark the point where computations of what is initially features of proximal stimuli generate representations of distal stimuli. For the idea is that this is also the point where distal properties become available and accessible to belief-forming processes (i.e. the point is the interface between the non-conceptual and the conceptual realms). So I take it that the claim Tye commits himself to by denying that their states are poised is that they aren't (at least not directly) sensitive to distal properties – lack of beliefs, as in his example with caterpillars, *indicates lack of access to external features*.

It should now become apparent that Tye is entitled to the response he gives provided it is determined unequivocally that blindsighters are sensitive to their surroundings by means of their motor reactions being guided by features of proximal stimuli. Justified as he may feel in moving from the absence of beliefs to the lack of distal representations even in the case of such anomalous phenomena as blindsight, a definitive answer is hardly to be provided on theoretical grounds. The question is just how great (indirect) sensitivity to one's surroundings proximal representations could support. For instance, what colour an object is perceived as having depends on what band of wavelength the light reflected by it is in, so an organism could be said, in one sense of the word, to detect colours while lacking spatial representations of colours completely. But there is a general consensus in the literature including (Tye 1996) (see, for example, Marr's influential (1983)) that detection of edges and surfaces requires computations over symbols which

represent beyond the properties of proximal inputs; such symbols appear in the latter stages of processing. And highly accurate estimation of shape and orientation, which - as highlighted by Carruthers in the above quotation - blindsighters are capable of, must presuppose access to shapes and surfaces as represented by their visual systems. It is therefore in this context that Carruthers' complaint that Tye simply overlooks how fine-grained their discriminations are seems particularly apt.

It turns out then that Tye must acknowledge that the undoubtedly non-qualitative states of blindsighters would qualify as qualitative ones on the PANIC theory in spite of their not lying at the interface between perceptual and belief-forming processing unless, to reiterate, the P is intended to imply that the latter processing renders the outputs of the former qualitative. Consequently, he will be compelled to attribute a range of wide contents where subjects are aware of seeing nothing at all.

This brings us back to the qualitative/non-qualitative division within the visual system the reality of which Carruthers argues for. What is preserved in blindsight is the non-qualitative route of processing where representations of external features are built and exploited in certain specific ways described above without acquiring phenomenality, without ever being displayed in 'the user interface'.

The interpretation of the experimental data on which Carruthers relies has been contested (see, for example, (Dulany, 2004) for alternatives), although it has to be said that its critics are in the minority. At any rate, we would be rightly inclined to suppose that perceiving without phenomenality may be widespread in nature. As it is clear from my presentation of the notion of poisedness, I take it to involve the following three claims: a) if the outputs of a visual system are representations of the properties of proximal stimuli, their link with actions will be brute-causal, b) phenomenal states would be of no use to creatures with no cognitive capacities (i.e. thoughts, beliefs, etc.), and c) if the outputs of a visual system are representations of distal stimuli, their link with actions is never brute-causal but is instead mediated by perceptual thoughts and beliefs. I have already indicated that I agree with b), for I believe that the only way phenomenal states can be exploited in a creature is by virtue of some aspects of their contents becoming contents of perceptual thoughts and beliefs. Neither is it reasonable to deny a), since it is hard to imagine how one could have qualitative representations of proximal stimuli;

patently, we do not see lightwaves and their properties such as wavelength and intensity. But c), in my view, is highly unlikely to hold for all types of visual systems which generate spatial representations. An organism's survival may depend on detecting the shape and orientation of objects but not on the ability to conceptualise perceptual contents. Such (non-qualitative) representations could still be causally efficacious in that system by virtue of their structure, that is, the transition from their being constructed to a behavioural output could be a matter of simple computations which are sensitive to what sets of symbols they are built from. Such transitions would certainly dispense with mental qualities. The situation is not dissimilar to that of, say, a photosensitive device which produces an electrical signal indicating the number of sides a polygonal figure placed in front of it has, or perhaps that of a bar code scanner. I struggle to find anything in Tye's account which convincingly rules out or at least undermines such possibility.

I conclude that there is nothing mysterious or contentious about the suggestion that there is a non-qualitative stream of visual processing in humans the existence of which is uncovered in numerous dissociation cases such as blindsight, and that one ought not to dismiss the interpretation of the experimental data to which Carruthers adheres as unintuitive, for an organism can be coherently imagined to exploit spatial visual representations to good effect even if they possess no phenomenal properties.

The argument so far has been that an appeal to Block's distinction between phenomenal and access consciousness fails to sustain FOR accounts due to the apparent difficulties in distinguishing qualitative perceptions from non-qualitative ones. Hence it will not do to say in response to the inattentive blindness objection that the absent-minded driver's experiences remain phenomenally conscious and that it is this type of consciousness that FOR is a theory of. Why is it much more fortunate to approach the problems of qualitative/non-qualitative perception and inattentive blindness from a higher-order perspective? I will now be concerned with the first one – seen, naturally, as the more fundamental one – and will then show how the treatment extends to inattentive blindness.

So we have, on the one hand, a first-order representational theory, on which the state that carries qualitative properties and the state that realises conscious (to mean phenomenally conscious) awareness are one and the same, and a higher-order

representational theory on the other, according to which they are two separate states, with the latter being a higher-order state aimed at the former. This difference between FOR and HOR is crucial for my purposes and it will be useful to introduce a distinction between qualities-bearing states and awareness-realising (again, intended to mean phenomenally conscious awareness) states lest we lose sight of it. First-order representationalists' contention that the two exist in unity is, of course, rooted in the thesis that mental qualities are represented properties - that they 'ain't in the head', as Putnam believed was true of the meanings of a certain type of linguistic expressions. It is then concluded that the (first-order) state which does the representing must simultaneously be an awareness-realising state.

In the preceding paragraphs attention was drawn to the possibility of non-qualitative spatial representations. They are genuine mental representations in the sense that they are involved in transitions which are sensitive to their contents. In particular, in the case of blindsight their contents have been found to play an essential role in guidance of fine-grained motor reactions. Proponents of FOR would be obliged to say that they are both qualia-bearing and awareness-realising even though, as an empirical matter, they are neither. Now, while, unlike FOR, HOR accounts do not merge the two into a single (first-order) state, both of them must nonetheless be present in order for conscious experience to arise, in other words, they both remain an integral part of conscious experience.

How does HOR accommodate non-qualitative perception? Firstly, it is important to appreciate the fact that it is open to proponents of HOR to claim that higher-order awareness-realising states cannot be tokened where the qualia-bearing component is missing. Blindsighters, of course, are not blind in their entire visual field and the suggestion does not apply beyond the region in which they are aware of seeing nothing. It is not the case that they are not subjects of higher-order states at all; rather, they do not enter higher-order states with respect to the perceptual intake in the blind region. I don't see how this idea can be disputed except on the grounds that neither purely recognitional higher-order concepts nor phenomenal concepts which are untranslatable into public language exist, but this will be the subject matter of the forthcoming sections where Sellars' story of how come to be aware of experiential inner episodes will be retold. Hence, these remarks provide the first clue as to the advantages which HOR brings in

comparison with FOR. Where advocates of FOR were found to be under a considerable strain in attempting to identify the difference between qualitative and non-qualitative perception drawing solely on the theoretical resources they are entitled to avail themselves to, the HOR theorist has identified a significant difference at a higher-order level. What is being proposed is that *a perceptual state is qualitative if it is available to be picked out as the content of a higher-order state.*

Before it is replied that it is unacceptable that a difference at the higher-order level could be of any importance as far as states at the first-order level and their phenomenality are concerned, let me contrast the kind of availability that is under discussion here with Tye's notion of poisedness. I pointed out that 'poisedness to make a difference in beliefs' is intended to mark the point where spatial representations are constructed, for, according to him, it is also the point where distal properties become available to belief-forming processes. Then I went on to say that if Tye resisted crediting blindsighters with PANICs despite their apparent (direct) sensitivity to distal stimuli, he would have to be interpreted as insisting that beliefs confer phenomenality on otherwise non-qualitative perceptions.

Since HOR too is a representational account, one might wonder whether the current proposal adds anything substantive to what Tye has already said about qualitative states, for if a state is poised to cause a first-order belief about its content, it will also be poised to cause a higher-order state about itself. The objection would therefore be that if I was reluctant to admit that blindsighters' states too would count as qualitative on this proposal, I could be understood as insisting that higher-order states confer phenomenality on otherwise non-qualitative perceptions.

But this rests on an oversimplification of HOR and it overlooks the crucial difference between HOR and FOR. Unlike Tye's poisedness, the kind of availability in question does not to establish a 'mere' functional link. In contrast to first-order perceptual beliefs, (higher-order) awareness-realising states are what FOR theorists thought was the same as the qualities-bearing state – they are an integral part of the architecture of phenomenal consciousness. We can also put this, as I did above, by saying that the (higher-order) awareness-realising state is what was held by FOR theorists to exist in unity, so to speak, with the qualities-bearing state. Although this, of course, cannot be taken literally – it is not that first-order representationalists discern an awareness-realising component and a

qualities-bearing component within the same (first-order) state - it still makes sense to say that HOR separates the former component from the latter and shifts it 'upstairs' to the second-order level. Thus the relation between the two within the HO picture of consciousness is far stronger than the functional relation qualitative states bear to first-order beliefs, for the suggestion is that when such relation obtains, conscious experience occurs.

The outcome of this discussion is, then, that HOR does succeed where FOR fails, that is, in marking the difference between qualitative and non-qualitative perceptions in a way which sheds much light on what it is for a state to be phenomenal. To say that phenomenal perceptions are those which are available to be targeted by higher-order states is to say something substantive (i.e. it does some genuine explanatory work) about mental qualities because the state to which it is so available realises, or constitutes, *conscious* awareness, where 'conscious' amounts to '*phenomenally* conscious'. And all this is claimed without being pushed into saying that the presence of higher-order states turns an otherwise non-qualitative state into a qualitative one.

In arranging the argument for HOR as I did I part ways with Carruthers in that he thinks that the kind of second-order contents which first-order perceptions yield on becoming available to higher-order states are also *perceptual* (or analog).¹⁷ But nothing in my argumentation prevents me from maintaining alongside FOR theorists that the only genuinely perceptual contents are the contents of first-order perceptions and, consequently, that the relevant higher-order state is a thought of the form 'I am in a state with such and such represented properties'.¹⁸ Despite this difference, I am prepared to embrace his distinction between worldly subjectivity and experiential subjectivity, as it expresses rather pithily what the advantage of any version of HOR over FOR (and indeed any first-order theory) consists in – it is in shifting the 'locus' of phenomenal consciousness to a higher-order level, where one, as he puts it, takes a subjective perspective towards his own experiences.

¹⁷ He does not object to being called a closet higher-order perception theorist after all, which is an approach also adopted by Lycan.

¹⁸ The issue of what kind of states they should be will be addressed at some length in the sections to follow.

This last argument substantiates the last two of the four claims listed at the beginning of this section. These were as follows: the claim (i) that there are qualitative states that occur in absence of consciousness, (ii) that there are states that are genuine instances of perceiving, to which the FOR theorist would have no choice but to attribute wide *perceptual* contents or face charges of inconsistency, but which nonetheless lack phenomenality altogether, (iii) that HOR has the resources to explain the difference between states such as those described in (i) and (ii) on the one hand, and conscious experiences on the other, and finally (iv), that only a HOR theory and not first-order representationalism is in a position to explain why my experience as of something yellow should feel like this rather than like something else.

With this HO picture of consciousness in place, I shall conclude this section by showing, briefly, how the treatment extends to inattentive blindness. This can be easily anticipated. We saw that the first-order representationalist's appeal to first-order beliefs, actions, or even verbal reports could not distinguish between the non-conscious experiences of the absent-minded driver and those accompanied by attention, since all of those were attributable to him while drove in absent-minded mode. What goes on when he switches to absent-minded mode is that, firstly, his (first-order) attention shifts from the contents of his perceptions to the contents of his memories, imaginations, etc., and, secondly, the perceptions themselves cease to be targeted by (non-conscious) higher-states. So there is no need to go down the precarious route and insist that his perceptions cease to be contentful states the moment he stops concentrating on his driving. They, however, retain their phenomenality precisely because they continue to be poised to cause higher-order states about themselves.

This completes the defence of HOR in this section. I shall now begin to set the stage for showing how Sellars' thought bears on the question of what kind of states realise higher-order awareness.

2.2 Higher-Order States as Internalised Public-Language Sentences

2.2.1 Subject-Dependence of Qualia

Lycan (1997) stresses that his higher-order sense theory makes no pretence of explaining qualia:

There may be Inner Sense theorists who believe that their views solve problems of qualia; I make no such claim, for I think qualia problems and the nature of conscious awareness are mutually independent and indeed have little to do with each other (p. 756).

Given the possibility of non-qualitative perceptions which represent properties of distal stimuli (their edges, orientation, etc.), it seems that Lycan is unduly modest about the scope of a higher-order theory. The insurmountable difficulty for the first-order representationalist brought out by Carruthers lay in his inability to rule out cases of perceptions where it was perfectly legitimate to ascribe wide contents by the criteria of his own theory while those states were devoid of phenomenal features. I added that even undermining the received reading of the blindsight data would be small consolation for the representationalist, since organisms which exploit non-qualitative perceptual representations of distal properties may be widespread in nature.

In the latter parts of the preceding section I outlined a certain pattern of higher-order explanation (of phenomenal as opposed to access consciousness) which successfully overcomes the difficulty. We may call it a ‘generalised pattern of higher-order explanation’ as it makes no assumptions about the nature of the relevant introspective states, that is, it remains neutral on the question of whether such states are themselves experiences or thoughts. The explanation proceeded as follows: a HOR theory separates out the awareness-realising component (to mean phenomenally conscious awareness), which was held by representationalists as well as all other first-order theorists to exist in unity with the qualia-bearing component within a first-order state, and shifts it to a higher-order level. The (first-order) representationalist does not of course discern the two

components within one state. He holds that qualia are contents and contents are outwardly directed. Hence a state which carries qualia as its contents also realises one's awareness of them (and since it is qualia that the state carries as its contents, their awareness amounts to phenomenally conscious awareness). He would feel that no explanatory ground would be gained by understanding these two features of the same state as discrete components, if any such understanding can be made intelligible at all.

Still, there is nothing unintelligible in stripping perceptual states of their (phenomenally conscious) awareness role (or feature) and relocating it to the second-order level where it becomes a component occurring alongside what has remained at the lower level – a qualia-bearing component. Note that once we operate within this HO picture and we have qualia as our explanatory target, that target is no longer what it was when approached from a first-order perspective. Having removed the awareness role, we are now dealing with the residue (earlier I called it 'residual qualia'). In other words, our claim is that states with qualia are less than phenomenally conscious when not complemented by introspective states, while the relevant introspective states do more than render the lower-order ones access conscious. So the generalised pattern progresses by disintegrating the original target of explanation.

It becomes apparent that the reason why Lycan believes that the scope of his inner sense theory does not reach beyond the conscious/non-conscious distinction (in the sense of access conscious) and that qualia problems are independent of this is that he leaves the original target unscathed, with the faculty of inner sense being an addition. FOR's appeal began to falter when the set of perceptions which satisfy its characterisation of qualitative states was found to include the kind of *contentful* states which do not require phenomenal properties in order to be utilised to good effect in an organism. The proprietor of such non-qualitative states is not precluded from displaying perceptual sensitivity to distal properties (i.e. shape, orientation, etc.), as opposed to the properties of proximal stimuli, if those states are symbolic structures whose symbols encode distal properties and appear in subsequent computations leading to a sufficiently accurate motor response in order for it to survive and thrive in a given environment. Let us label a representational system whose sensory apparatus, as well as that system as a whole, dispenses with mental qualities an 'unconscious perceiver'.

Suppose that, as my generalised pattern of higher-order explanation recommends, the question of why certain contentful perceptions are qualitative while other equally contentful perceptions aren't reduces to the question of why a certain class of representations have residual qualia while others don't. The pattern will differentiate between them on the basis of availability to introspection (or higher-order awareness-realising states). Given that the absent-minded driver happens to be blessed with the relevant capacity to introspect, his unheeded perceptions will stand ready to attract introspective attention to themselves, thereby retaining their residual qualia. On the other hand, the unconscious perceptual systems in the example given two paragraphs above would lack the requisite higher-order states, meaning that their perceptions would not be so available. Consequently, their perceptions would be bereft of residual qualia.¹⁹

But what line of thought reveals that all qualia are to be treated as residual and that the awareness-realising role should be transferred to the upper level? Is such treatment not a matter of stipulation which secures that my pattern of explanation differentiates between us and unconscious perceivers but is otherwise unsupported? Furthermore, how can an account on which qualia (or residual qualia) are construed in terms of their availability to introspection be a substantive explanation of qualia?

I believe the desired reinforcement is supplied by certain reflections on subjectivity. Dennett's observations in his (1978) are a useful starting point:

Functionalist theories are theories of what I have called the sub-personal level. Sub-personal theories proceed by analysing a person into an organization of subsystems (organs, routines, nerves, faculties, components – even atoms) and attempting to explain the behaviour of the whole person as the outcome of the interaction of these subsystems (p. 153).

¹⁹ Carruthers would here encourage us to bear in mind the distinction between worldly and experiential subjectivity. According to him, blindsighters' states possess the former, but lack the latter type of subjectivity, which, again, is an ability to "take a subjective perspective towards one's own experiences" (2000). Therefore, '*phenomenal* consciousness' must stand for experiential subjectivity rather than worldly subjectivity. However helpful Carruthers' worldly/experiential distinction with respect to subjectivity is in making intelligible the claim that my generalised pattern of higher-order explanation is at least an approximation to a theory of qualitative properties, there are significant differences between my proposal and his dispositionalist higher-order thought view. The differences will be brought into focus below.

And several lines below he adds that ... “they would characterise relations not between a person and a body, or a person and a state of affairs or a person and anything at all, but rather, at best, relations between parts of persons (or their bodies) and other things” (p.153). In this last sentence he talks about relations characterised by his notions of computational access *simpliciter*, which is the kind of access the executive unit in a computing machine has to the outputs of its subroutines, and computational access of a print-out faculty which is analogous to access afforded by verbal reports of lower-order states in humans. He then embarks on the project of carving “a full-fledged ‘I’” – the experiencing self - out of these notions. Notwithstanding the peculiarities of Dennett’s project, his remarks that the totality of sub-personal relations seemingly omits the experiencing self extend to all versions of functionalism, with the intuitive consequence that

At best a sub-personal theory will seem to give us *no grounds* for believing its instantiations would be subjects of experience, and at worst (as we have seen) a sub-personal theory will seem to permit instantiations that *obviously are not* subjects of experience (p.154, italics original).

I do not wish to assess functionalism’s prospects of harnessing the self, although at this stage I do not see any reasons why such accounts should not be worth pursuing. However, I take Dennett’s observations in these passages to show that functionalist approaches which do not even pursue the self with whatever resources they have available can make no legitimate claim to have an answer to the problem of qualia. Theories of qualia which leave the questions to do with the self-reflexive ‘I’ which grounds higher-order (introspective) states aside are typically those on which such questions are unrelated to qualia (Tye’s and Dretske’s positions are an example of this, whereas Lycan does give a theory of introspective access, but thinks it is irrelevant to qualia); by doing so, I will argue, they misconceive what is unique to experiences and what sets us apart from unconscious perceivers and simple photosensitive devices. And if their negligence of the personal level does lead to a misconception, then qualia will

inevitably be residual, with the higher-order level grounding *phenomenally* conscious awareness.

Lycan and Carruthers are among the functionalists who do set out to give an account of the self and join Dennett in his belief that it must somehow precipitate out of sub-personal parts. Lycan's organ of inner sense, as well as Carruthers's mind-reading module, is just one of those sub-personal parts listed by Dennett. Take, for instance, Lycan's inner sense. Saying "My inner sense is in me" would appear as innocuous as saying "My kidney is in me". Nor does "My inner sense is in me" entail "I am in me", let alone that all of the functions which the organ performs are realised in one particular corner of the brain with clearly defined boundaries to the effect that if it was surgically removed I would cease to be a person and would be legally treated as a population of cells. The realisers of its functions may be scattered around the entire brain.

But when one's description of how the organ interacts with other modules, routines, etc. is sufficiently exhaustive, we can be hopeful that we have at least an approximation to a theory of the personal level, despite the fact that the organ continues to be describable, from a certain perspective, as a sub-personal part.²⁰ A functional nature of the former description ensures that the above entailment does not obtain. "My inner sense is in me" is embedded in the latter one, under which it is a sub-personal part, while the way the organ *functions* renders "I am my inner sense" false, i.e. it functions in such a way which guarantees that I take myself to be what I pre-theoretically took myself to be – and I certainly never believed I was an inner organ in this body. But of course the very challenge for these functionalist conceptions is to come up with descriptions of causal interactions, whether of the inner sense or the computational access of a print-out faculty or else, which mirror the logical behaviour of the first-person pronoun (see John Perry (1988) and Castañeda (1966) for an extensive treatment of so-called *de se* ascriptions of knowledge).

Having made these points, I can begin to formulate my argument for residual qualia. First-order representationalists' portrayal of phenomenal properties makes no mention of

²⁰ That we can have a description under which the organ is a sub-personal part alongside one under which its interactions with other modules constitute the self is of course a commonly appreciated advantage that functionalism brings in contrast with identity theories. Such result would be unattainable if the truth of an identity theory was assumed.

a self-reflexive capacity and the higher-order states it facilitates. Plainly, the account being first-order, we should expect it to reach its explanatory target while staying within first-order confines. Moreover, it is not difficult to see that virtually any such mention would also threaten its bare existence. Even inclusions of fairly flimsy connections with introspective states would amount to making concessions to HOR's proponents, thereby gradually heading towards abandoning the theory that the representationalist has under construction. For some HOR theories, such as Carruthers', are dispositionalist rather than actualist and do not demand that a higher-order (introspective) state be actually present in order for a higher-order *content* to be created in an experiencer whereby he becomes introspectively aware of his perceptions. A dispositionalist would only insist on as little as a relatively tenuous functional connection with the mind-reading module. Take the example with carburettor, often used to test basic functionalist intuitions. A carburettor does not cease to be a carburettor when the engine stops. It remains one for as long as it is connected to certain other units in the right way and does what it was designed to do when the engine starts. Those other units need not be in operation in order for a carburettor to 'persist', just as the mind-reading faculty need not be in operation in order for higher-order contents to persist.

That the self-reflexive 'I' can barely play even a minor role in FOR's characterisation of mental qualities is its major deficiency. This is because the omission of the self deprives the representationalist of any means of moving from "There is a visual representation of something red in me" to "*I* visually represent something red". Consider the two statements as alternative descriptions of my current experiential situation in which I find myself by being presented with a red object. On representationalism, my perception is a qualitative one provided the former description is satisfied and applies to my experiential situation. The representationalist would of course add that only representations of properties of external objects (rather than of properties of proximal stimuli) qualify as qualitative, so let the term 'representation' as it appears in the description refer exclusively to such representations. However, only the latter description, which can be attributed to HOR, entails that *I* am visually conscious of something red. It is a basic fact about phenomenal states – at least if the truth of the claim that perceptions are intentional is assumed – that they make their subjects (non-conceptually) conscious of

what they represent. On the other hand, “There is a visual representation of something red in me” by no means entails that my perception performs its representing function for me to the effect that I can be described as “I visually represent something red” or “I see something red”. It entails, at best, that I am an *observer* of my perceptions rather than their *subject*.

The other, and more serious, consequence of the failure of the desired entailment in the case of FOR is somewhat harder to bring out. The following is my best effort. We are often tempted to think of qualitative properties as mental paint, and this is perhaps innocuous, if only as part of some pre-theoretic rumination. We then go on to ask ourselves how our states can acquire mental paint. If I’m right in my suspicion that the representationalist puts us in a position of observers rather than subjects of our sensory states, he will tacitly allow the pre-theoretic inclination into his theory of qualia. He will of course insist that it only makes sense to speak of mental paint in so far as it is construed as content. But, contrary to his insistence, it appears that something which has phenomenal properties, but lacks a proper subject and can only be observed instead comes curiously close to what phenomenologists wanted to say about mental paint. It is as though I, or, for the sake of illustration, my ‘inner eye’ stands at a distance from my representation observing the representation’s ‘mental paint’, and I proclaim, having representationalism as my background theory, that that ‘mental paint’ is in fact its intentional content. In other words, it is as though ‘mental paint’ is already present in its full richness as a lower-order representational content and introspecting it consists in merely exposing one’s inner eye to it.

Such picture is fallacious and ought to be discarded. I do not stand at a distance from my (qualitative) representation of something red, for what really takes place is that I visually represent something red. I, or my *self*, is submerged, figuratively speaking, in that ‘mental paint’.

So my objection is that first-order views of all varieties preserve the pre-theoretic tendency to believe in mental paint. “There is a visual representation of something red in me” implies that I am related to phenomenal properties in a way analogous to someone scrutinising a painting in that phenomenal states can exist in representational systems where nothing can be properly identified as the subject of those states, just as real paint

exists regardless of anyone scrutinising it. The representationalist should find such implication disconcerting. On the one hand, he will be eager to hold on to the claim that qualia are intentional and that the two cases are disanalogous in this regard, since it is absurd to say that real paint is a painting's intentional property, but on the other, having given the self the role of an 'external' observer in relation to qualia, he accords qualia the kind of objective and independent status which only a phenomenalist would be comfortable with. As a relationalist, he would surely resent any association with phenomenism. Phenomenalists take it that qualitative properties are not the product of one's perceptual contact with (external) objects which are by no means literally coloured, rather, they themselves are sensed. On this naïve realist view, sensing is an act of exposing oneself, or one's senses, to what is phenomenal in its very essence prior to that act. We now have every reason to suspect that prising the self from qualitative properties puts an initially relationalist project at risk of being assimilated to a naïve conception of qualia, the claim that qualia are intentional losing all significance.

Unless we are naïve realists about phenomenal properties, we should not accept that they can be found in creatures who are not subjects (or selves), to whom the description "I visually represent something red" does not apply. The fact that the personal level is integral to mental qualities in the way just outlined is the primary reason that they resist reduction more than many other natural phenomena. It is not that the phenomenal redness that I have in front of me has an objective status in the sense that I can coherently wonder whether, for example, it has a mass, and that my enquiry is similar to that of a physicist who wonders whether photons have a mass. The redness is a subject-dependent property and our enquiry must start at the subjective or personal level if we are to arrive at an understanding of why it appears to me that there is a patch of redness with this particular phenomenal 'feel'. Some naïve realists may of course object that phenomenal properties are mind-dependent. But I see no obvious move from mind-dependence to subject-dependence.

Further, unless we aspire to give an account of the personal level, whether we seek to build it from sub-personal material or otherwise, we will not move beyond giving an account of such representational systems as unconscious perceivers and photosensitive devices. I do not wish to deny that our experiential situation continues to be describable

as “There is a visual representation of red in me”, just as Lycan’s inner sense continues to be describable, from a certain perspective, as a sub-personal part even if it is granted that facts about its causal interactions with other modules and subroutines exhaust facts about persons; for this much (i.e. “There is a visual representation of red in me”) we have in common with unconscious perceivers who replicate our first-order functional organisation. But as soon as the observation that my self is woven, so to speak, into my perceptual representation’s fabric receives full appreciation, only facts about the representation’s interactions with the relevant self-reflexive states such as “I see something red” or “I visually represent something red” can begin to explain why it is a qualia-bearing state with this particular ‘feel’. Therefore, again, qualitative states cannot occur where descriptions such as “I visually represent something red” do not apply. They cannot occur in creatures which are not endowed with a faculty of introspection.

My generalised model of higher-order explanation accommodates the extent of qualia’s dependence on one’s ability to introspectively focus on his perceptions by characterising qualitative states in terms of availability to introspection. Now that it seems undeniable that the self is inextricably involved in experience, we should be prepared to acknowledge that the model’s appeal to availability (together with its representationalism about residual qualia) does express something substantive about what makes a certain class of sensory representations qualitative and that the model serves at least as an approximation to a theory of qualia – provided it is underwritten by a fully developed account of the personal level.

The availability relation of course requires further elucidation and it will be the subject of a more extensive discussion in chapter 3. Here I restrict myself to making the crucial point. It is instructive to return to inattentive blindness again. As I drive absent-mindedly, I stop at traffic lights. It is uncontroversial to say that the perceptual representation whose content I absent-mindedly act on is a qualitative one. The point to emphasise is that the description “I visually represent something red” is true of (or applies to) my present situation irrespective of whether or not I attend to the perception. It applies because I’m the *subject* of the state – and to say that I’m the subject of the state is to say that I can, should the need arise, *form* a higher-order thought about it, one that is equivalent to the theoretical description “I visually represent something red”, such as “I see something

red” or “I am visually conscious of something red”. The theoretical description can be true of situations in which introspection lies idle. Someone who is already sceptical about the soundness of my argument might point out that an appeal to introspection does not in itself differentiate between my model of explanation and the picture of qualia as mental paint that I criticise. The kind of access the self understood as an observer has to qualitative properties also qualifies as introspective after all. But there is an important difference between introspecting my state as the subject of that state and introspecting it as its observer. It is the former that I appeal to.

My argument against FOR is also a robust justification for treating qualia as residual. Since only “I visually represent something red” can be an adequate description of our experiential situation, and the first-person pronoun it contains is a product of a self-reflexive faculty, it appears natural that the awareness-realising component should be found at the higher-order level, meaning that a state’s possessing qualia is a necessary but not sufficient condition of phenomenal consciousness. That their presence is only a necessary condition is of course all I had in mind when I called them ‘residual’. Incidentally, I believe that an approximation is as much as the model ultimately has to offer. It points us, however, in the right direction. In the final chapter I will attempt to make my own proposal concerning the nature of phenomenal properties, one which escapes the rather far-reaching consequences of the issue I raise in this work. We will see that my proposal draws much of its strength from the model currently under discussion. In addition, the problems (to be discussed in the next two sub-sections) the model, as well as all other HOR theories of qualia, encounter and are ill-equipped to deal with provide some vital insights to be taken account of when formulating my alternative proposal.

So the upshot, to reiterate, was that representationalism fails to identify the subject of experience, as indeed all first-order theories do, and is therefore vulnerable to counterexamples involving representational systems which realise our first-order functional organisation but experience nothing at all. And if one continues to insist that phenomenality can be found in them despite their not being subjects (despite knowledge *de se* not being ascribable to them under any circumstances), he will face the charge of being a closet phenomenalist postulating subject-*independent* mental paint.

As for HOR, devising a suitable counterexample in which an imaginary system also replicates our higher-order functional organisation would be a more laborious task. But Dennett's worry was that in principle the task is never an impossible one if the theory in question is a functionalist enterprise building the personal out of sub-personal material; HOR theories are functionalist in their core. After all, computers can scan themselves, search for damaged files, or sort stored images according to their various features, without becoming subjects of experience.

This, finally, brings us the topic of recognitional concepts. A proponent of HOR may try to dispel Dennett's worry by suggesting that introspective thoughts directed at experiences embed a special kind of concepts – so-called phenomenal concepts which are essentially recognitional. As I will explain at some length below, phenomenal concepts are recognitional in the sense that they are acquired solely by means of a subject's acquaintance with his experiences and their phenomenal properties, and so are grasped independently of any public-language or language of thought concepts. Their independence permits, or even presupposes, diversification of the self into the experiencing self and the conceptualising self – the experiencing self deploys phenomenal concepts, thereby allowing for the possibility that pre-linguistic children too are such selves, whereas the conceptualising self forms concepts whose reference and meaning are at least partly determined by their inferential roles and emerges exclusively when one acquires such concepts. Inferential roles can of course be understood computationally (in some Fodorian sense (1975)) or normatively (i.e. prescriptively – Sellars (1997), Brandom (1998), Gibbard (1994), etc.), in which case the conceptualising self emerges either when one acquires concepts belonging to the language of thought, or when he learns a public language. But this is a marginal issue in the present context as either alternative contrasts sharply with what is supposed to be true of concepts belonging to the experiential domain – that their reference and meaning is fully determined by causal relations with their referents; and if they do begin to figure in inferential transitions on one's acquisition of a language, their inferential roles make no contribution to determining their meaning and reference.

Thus a proponent of HOR seemingly regains his footing by introducing recognitional concepts into the HO picture of phenomenal consciousness. The presence of the

experiencing self together with its unique set of concepts suffices for the description “I visually represent something red” to apply and distinguishes sentient selves from unconscious representational systems equipped with self-scanning modules. The approach has the virtue of narrowing our explanatory target by not demanding that sentient creatures also be sapient. Sapience and rationality, which are the characteristics of conceptualising selves, are deemed irrelevant to what we share with small children and many higher animals – a somewhat more primordial form of subjectivity. However, after making a few introductory remarks about recognitional concepts of experience in the next sub-section, I will exploit Sellars’ distinction between observational and theoretical discourse to argue that recognitional concepts are an epistemologist’s fiction; a story of how we *actually* come to form a conception of our experiential states makes no room for recognitional concepts. I will then hope to show that the HO structure of phenomenal consciousness collapses when unsupported by phenomenal concepts (I use the terms ‘phenomenal’ and ‘recognitional’ with regard to our concepts of experiential states interchangeably).

Before I move on, let me indicate briefly which stage in my overall argument we now find ourselves at. The argument was made up of the following inconsistent set of claims:

- (i) Qualia are not intrinsic properties of perceptual states, and if so, they must be wide contents.
- (ii) The thesis that qualia are wide contents can be plausibly held only as a claim of higher- rather than first-order intentionalism.
- (iii) To have a higher-order thought about one’s perceptual state is to token an internalised public-language sentence whose meaning is holistic and normative.
- (iv) Since higher-order perceptual thoughts are internalised public-language sentences, qualia are intrinsic properties of experiences.

In chapter I. I defended (i) against certain rival views of perceptual content and its relation to qualia. I made no attempt at a full defence as the claim is intended to serve mainly as a working assumption for the sake of my argument. However, a full defence of (ii) was my objective in the first section of chapter II. as well as the present section up

until this point. A sellarsian-style argument of the sort presented below will inevitably lead to (iii). In the final section of this chapter I will undertake to show that (ii) and (iii) together result in the claim that qualia are intrinsic properties of experiences – a consequence expressed in (iv).

I shall now focus on the subject of recognitional concepts.

2.2.2 Recognitional Concepts

Recognitional concepts of phenomenal properties are not a higher-order representationalist's invention. The proposal that we possess such concepts traces its origin to the dispute concerning the metaphysical consequences of the so-called conceivability argument formulated by Jackson (1982 and 1996), and Chalmers (1996). While Jackson and Chalmers hasten to draw firm ontological conclusions from conceptual possibility of absent qualia, a number of authors (Levine (1993) and (1998), Loar (1990 and 1996), Papineau (1993a) and (1993b), Sturgeon (1994 and 2000), Tye (1999), for example) have been more than reluctant to attach any ontological import to it. The source of the notorious explanatory gap between phenomenal properties and physical/functional properties is, according to them, the unique nature of the concepts we deploy to refer to phenomenal states as their subjects. Thus recognitional concepts of phenomenal properties are introduced as a means of blocking the inference from conceivability (i.e. conceptual possibility) of absent and inverted qualia to metaphysical possibility of absent and inverted qualia and sustain the gap only in some weak epistemological sense.

The term 'recognitional' is only used by Brian Loar, but the essential characteristics of experiential concepts that the term is to help make explicit are, as we shall see shortly, also highlighted under various guises in the works of the other authors listed above. In order to get a grip on these essential characteristics, we need to understand how they undermine the conceivability argument. To begin, Levine (1998) notes in his lucid exposition of the argument that it relies on what he calls 'the distinct property model' (DPM) of explaining "a priori ignorance of an identity" (p. 455). These are cases of a

posteriori identities such as “Water is H₂O” the truth of which one may fail to realise despite possessing both of the concepts appearing in the identity claim.

The following is the line of reasoning that the model would licence: On a certain (rather traditional) conception of the semantic properties of concepts, concepts have an extension (or a reference) and a mode of presentation. Mode of presentation is also referred to in the literature as ‘intension’, ‘connotation’ or ‘inferential role’ (whether computational or normative). The intension of ‘water’ is expressed along the lines of “transparent, odourless liquid, etc. which fills rivers and lakes” but the details may vary from thinker to thinker depending on what inferential role the concept ‘water’ has in them. The mode of presentation of ‘water’, whatever its precise details are in a thinker, is only contingently related to the mode of presentation of ‘H₂O’. If it was necessarily related, the identity “Water is H₂O” would be knowable a priori; surely, scientific identities are established as empirical discoveries, not as outcomes of a priori analyses, which obviates how an identity can be unknown to one despite his having the concepts of properties that have been successfully identified within a certain field of science. And, Levine continues, if the point about contingency is combined with an internalist view of meaning (assumed by the DPM), on which the sum of observable properties expressed in the mode of presentation of ‘water’ exhausts the meaning of the concept, then “Water is H₂O” can at best be a *nomologically* necessary identity. For although “transparent, odourless liquid, etc. which fills rivers and lakes” would pick out H₂O in our world (considered as actual), it would pick out XYZ on Twin Earth (a world considered in this context as counterfactual). Hence the inference from conceptual possibility (i.e. conceivability) of water not being H₂O to metaphysical possibility of water not being H₂O.

It was an insight of Kripke’s that proven scientific identifications are metaphysically necessary and that ‘water’ refers to H₂O in every possible world in which it is a referring expression at all and it is now common ground between materialists and anti-materialists. Jackson and Chalmers do not of course question metaphysical necessity of “water is H₂O” and the above is only an illustration. Their claim is that the inference goes through for pairs of properties such as phenomenal property R (red) - physical/functional property P. It is identifications of these that can at best be nomologically necessary. Since the

mode of presentation of 'phenomenal property R' is only contingently related to the mode of presentation of 'physical/functional property P', it follows from conceivability of R not being P that it is metaphysically possible that R is not P. "Water is H₂O" allegedly differs from "R is P" in that a description of the observable properties of water can be formally derived from the relevant chemical description (of H₂O) if the non-rigid reference-fixer of 'water' in the above example is rigidified as what Chalmers and others call 'primary intension', associated (in his (2006)) with 'fregean' intensions. Block and Stalnaker (1999), however, reveal serious flaws in employing derivability as a measure of identification's success. Moreover, given Jackson's and Chalmers' method of isolating the primary intension of a concept – i.e. as a function from centred worlds to extensions – they quite rightly see no reasons why primary intensions with a similar degree of rigidity could not be assigned to concepts referring to phenomenal properties to the effect that they would designate the same physical/functional properties in every (centred) possible world, thereby meeting the criterion of formal derivability. Jackson and Chalmers, they conclude, fall short of showing an asymmetry between "Water is H₂O" and identities such as "R is P".

Levine does not attempt to undermine the conceivability argument by disputing whether the primary intension of 'phenomenal property R' is any less rigid than the primary intensions of standard natural kind concepts. Instead, he considers dropping altogether the internalist assumption about meaning in the DPM on which the argument is heavily reliant. Neither primary intensions nor modes of presentation as they are traditionally understood make any contribution to the meaning of introspective concepts of phenomenal properties. He states this by saying that such concepts have non-ascriptive modes of presentation:

An ascriptive mode is one that involves the ascription of properties to the referent, and it's (at least partly) by virtue of its instantiation of these properties that the object (or property) is the referent. A non-ascriptive mode is one that reaches its target, establishes a referential relation, by some other method. The object isn't referred to by virtue of its satisfaction of any conditions explicitly represented in the mode of representation, but rather by standing in some particular relation to

the representation. The mode of presentation is the relation itself. The usual candidate for such a relation is some causal or nomic relation, such as covariation between representation and referent that meets certain constraints (1998, p. 457).

If the concepts in question do have non-ascriptive modes of presentation, drawing any ontological conclusions from conceivability of absent and inverted qualia becomes a far more troublesome business. My concept 'red' as I apply it introspectively may refer to a physical/functional property without this fact ever showing up in my set of introspective beliefs about the property and its distinctive 'feel', in other words, without my introspection supplying any information about the physical/functional state my brain is in at the time of my having the sensation. Likewise, my concept 'water' designates H₂O without this fact necessarily showing up in my corresponding narrow psychological state, i.e. in my grasp of the concept. The inferential role of my introspective concept 'phenomenal property R' differs profoundly from the inferential role of 'physical/functional property P', and yet this does not preclude the statement "R is P" from expressing a *metaphysically* necessary truth.

We can see already from the brief outline in the quoted text the connection between non-ascriptive modes of presentation and Loar's phenomenal concepts which are recognitional in that they are acquired by means of one's acquaintance with phenomenal properties. The following is just some evidence of the parallels:

A recognitional concept may involve the ability to class together, to discriminate, things that have a given objective property. Say that if a recognitional concept is related thus to a property, the property triggers applications of the concept. Then the property that triggers the concept is the semantic value or reference of the concepts; the concept *directly* refers to the property, *unmediated by a higher order reference-fixer* (Loar 1990, p. 87, italics mine).

"Directly refers to the property, unmediated by a higher order reference-fixer" is to say that a phenomenal concept does not attach to its referent "by virtue of its satisfaction of any conditions explicitly represented in the mode of representation" (previous quotation).

If we are to state this in terms of the two-dimensional intensional framework, we may say that the meanings of phenomenal concepts are fully determined by their secondary intensions, ones which are isolated as functions from uncentred possible worlds to extensions and which Chalmers associates with ‘russellian’ intensions.

So what we have here is a radical form of externalism with clear atomistic implications, recognising of course that if it is put forward exclusively for concepts of qualitative properties, it will be a theory of referential relations to properties which are *internal* to the subject. We should perhaps call it ‘extensionalism’ instead to avoid the superficial contradiction of terms and to place emphasis on the fact that, at least as far as concepts of qualia are concerned, it equates meaning with reference. Nothing with atomistic implications is defended by Putnam (1975) or Burge (1979); their version of externalism is reconcilable with meaning holism as well as molecularism (Putnam’s (1991), for example, is a blend of externalism and normative holism) – the meaning of a natural kind concept is a result of a joint contribution of contextual factors and inferential roles.

There is a further motivation for the idea that concepts of qualia refer directly, without the aid of a reference-fixing mediator, besides its merits in weakening the conceivability argument. The point is a familiar one but it is particularly illuminating in this context to borrow Sturgeon’s terminology (2000) to bring it out. A concept is canonically linked to evidence for its application if the evidence is (at least partly) individuating of the concept. Canonical evidence acts as an *intermediary* between a concept and the property it applies to. Further, where there is an evidential intermediary, there are conservative defeaters, meaning that canonical evidence is always in principle defeasible; a concept can be misapplied even when canonical evidence for its application is available. Sturgeon gives this example: Let ‘red’ be a concept which refers to properties of external objects rather than to properties of experiences. My experience with a red quale acts as an evidential intermediary between my concept ‘red’ and the (external) property referred to, and the evidence is canonical (i.e. at least partly individuating of the concepts), as ‘red’ would not be the concept it is if it wasn’t warranted by such experiences at least in certain paradigmatic cases of its application. The evidence is defeated when an object of a different colour, due to unusual lighting, produces a red sensation in me.

The terminology can be used to highlight a number of epistemic and semantic characteristics of ‘water’ in a similar fashion. A set of observational properties such as being colourless, odourless, liquid, etc., is an evidential intermediary between my concept ‘water’ and the property it refers to. The evidence is in principle defeasible – a substance with all of the properties which make my central beliefs about water true (suppose, for simplicity, that I have no knowledge of chemistry) could nonetheless fail to be water (i.e. H₂O), although ‘water’ would not be the concept it is if it wasn’t warranted by the presence of these properties at least in some standard cases of its application. The defeasibility of such evidence stems from the fact that ‘water’ designates rigidly and invariably picks out H₂O. ‘Red’ too, as it appears in ordinary usage, is a natural kind concept after all. It denotes a physical kind – a kind of surface reflectance producing a range of states which their subjects classify as red. The surface-reflectance property of objects in question is, as the referent of ‘red’, stable across all users of the concept. This ought to be conceded even by those who take the possibility of qualia inversions with all seriousness – even if the state on the basis of which I apply it differs qualitatively from the state which normally warrants someone else’s application of it under the same (standard) circumstances.

There is a notably strong correlation between a concept’s canonical evidence, understood as a defeasible intermediary, and its contingent reference-fixing mediator. The property the presence of which I take to be evidence for applying a concept is also mentioned in the description that expresses the mode of presentation of the concept and states, at the same time, its contingent reference-fixing condition. If the presence of a red sensation is the canonical evidence for ‘red’, then ‘the cause of red sensations’ expresses the concept’s mode of presentation and states its contingent reference-fixing condition. Plainly, what one takes to be evidence for a concept’s application is part and parcel of his grasp of the concept, of its mode of presentation.

The correlation between the two is also apparent from the fact that if canonical evidence for applying a concept does not act as a (defeasible) *intermediary* between the concept and the denoted property, the concept refers directly, unmediated by a contingent reference-fixer. Absence of an evidential intermediary indicates absence of a contingent

reference-fixer. Qualitative concepts, unlike 'red' and 'water' in the previous examples lack evidential intermediaries:

Introspection disallows conservative defeaters. ...when I introspectively notice one of my mental states is q-red, my canonical evidence for this isn't some evidential intermediary between the q-redness of my mental state and my introspection-based belief about that q-redness. My evidence just *is* the q-redness of the mental state. Introspection moves directly from the phenomenon in question to our conception of it. This is why it feels immediate (Sturgeon 2000, p.48, italics original).

Sturgeon does not associate this fairly trite point about introspective concepts with talk of contingent reference-fixing conditions. But it is nonetheless easy to see how it motivates the view that there is nothing in one's grasp of a qualitative concept which makes reference to a contingent reference-fixing description. The possession-conditions for any introspective concept make reference solely to the causal relation between the concept and the property that triggers its application, since introspection is not defeasible and "moves directly from the phenomenon in question to our conception of it".

It seems that these rather trivial remarks about introspection are the key source of support for the idea that we have recognitional concepts of experiences. And many would of course feel that no additional support is required – only those approaches which view concepts of qualia as being formed on the basis of one's direct acquaintance with their referents, without the aid of higher-order reference fixers, can do justice to the kind of immediacy peculiar to our knowledge of mental qualities. I will argue shortly that the same impression of introspective immediacy can be created by reconstructing our conception of experiential states, as Sellars does, out of certain objective and intersubjective elements. Moreover, Sellars' argument that experiences and their properties initially play the role of theoretical postulates within a person's conceptual scheme, and enter into observational discourse only after a further enrichment of the

scheme provides overwhelming reasons to favour such reconstruction. Recognitional concepts will then turn out to be nothing more than an epistemological fiction.

2.2.3 HOR's Commitment to Recognitional Concepts

Contextual determinants of referential relations go undetected by users of concepts for which externalism was proposed. No amount of scrutiny of the internal component of meaning uncovers any traces of what lies at the outer end of a referential relation. This, after all, is why the position is known as 'externalism'; if contextual factors could be gleaned from the purely narrow component of meaning and were somehow implicit in it (whether 'narrow' stands for 'inferential role', 'intension', 'mode of presentation' or other), then that component alone would surely individuate concepts finely enough to do justice to modal intuitions involving superficially matching counterfactual worlds with different chemistry. If this was the case, no need for an externalist alternative would ever arise.

There has been much bewilderment among authors as to how a concept's referent can be concealed from its competent users that led to objections that externalism is incompatible with privileged access (see McLaughlin and Tye (1998) for possible responses). In sub-section 1.1.2 I evaluated the argument from incompatibility in the form it takes in Searle's (1983) regarding perceptual content. Searle swiftly dismisses externalism on the grounds that external factors do not enter into a perception's intension and therefore, assuming that only intensions are psychologically relevant, it has no role to play in individuating content.²¹ But the key insight of externalism is that external determinants and intensions are virtually independent of each other, that the former are *ex-ternal* to the latter while being individuating of content, not that they should be discernible within the narrow component. Searle's misgivings about contextual factors not entering into intensions place an impossible demand on externalism; if intensions did incorporate contextual factors, the resulting position would remain internalistic, and so

²¹ The belief that only narrow content is relevant to psychological generalizations is subjected to compelling criticism in Burge's well-known paper *Individualism and Psychology* (1986).

what he is in effect demanding, somewhat paradoxically, is that externalism can be true only if internalism is true.

Externalism's alleged clash with intuition stems from the widely shared conviction that we, as competent speakers, ought to be accorded the ultimate authority in decisions on what our concepts designate. The conviction becomes still deeper in the case of qualitative concepts which, being a species of introspective concepts, apply to items which are proprietary to the subject and are themselves thought to be known with absolute or near-absolute authority, in other words, we do not just have direct and infallible knowledge of the internal or a priori component of introspective concepts, as it is believed to be the case with all other concepts, we also have such knowledge of the essence of what they apply to. So if privileged and immediate access and undetectability of contextual determinants from the speaker's perspective are seen as two conflicting ideas as far as the referents of concepts such as 'water', 'gold', 'mammal', etc. are concerned, we should expect any attempt to reconcile them in a theory of qualitative concepts to be rejected out of hand. And many authors with anti-materialistic sympathies have indeed been scornful of the suggestion that the true nature and ontological status of the referents of qualitative concepts is concealed from their possessors – that they may designate physical/functional properties without their possessors having any knowledge of this fact from inspecting their experiences.

However, the purpose of my connecting Sturgeon's observations about introspection with Loar's thought that qualitative concepts do not designate via contingent reference-fixers was to show that, in the case of qualitative concepts, not only does the idea of immediacy cohere with undetectability of contextual determinants, but it is also the primary source of motivation for radical form externalism, or 'extensionalism', regarding these concepts. Introspection feels immediate because, as Sturgeon writes, canonical evidence for applying a qualitative concept is not an intermediary between the concept and the property it applies to, rather, it is the phenomenal property itself and therefore, in introspecting, we move directly from the phenomenon in question to our conception of it. This is arguably akin to the sense of 'immediacy' to which anti-materialists would appeal.

Now, if my concept of, say, a purple quale is formed in such a direct fashion, it follows that the sole constraint on my possession of it is that I stand in an appropriate causal relation to the property. All of the reference-fixing that is necessary in order for that concept to be the concept it is, with its specific degree of abstractness or determinateness, the range of shades it covers, etc., is done by the causal relation, leaving no room for a narrow and contingent description to share, so to speak, some of the reference-fixing tasks; and since there is nothing in my manner of possession of it except that causal contact, the concept is devoid of an analysable purely internal or a priori component.

This deprives Searle and others of any grounds for complaining that contextual factors do not enter into intensions. For, owing to the absence of an a priori component, there is nothing for them to enter into; or, alternatively, there is nothing – no intensions, that is – they are *ex*-ternal to, except the concept itself. Bearing in mind the extent of the conceivability argument's reliance on internalist semantics, as it was clear from Levine's exposition, it also deprives anti-materialists of any means of arguing from cognitive independence of qualitative and physical/functional conceptions to ontological distinctness of their referents. The narrow components having been removed, there is nothing a priori left in phenomenal conceptions to dictate that they cannot share their extensions with physical/functional ones.

A corollary of a treatment of qualitative concepts which abandons all internalist assumptions about their meaning is that it inevitably severs whatever a priori links we might have supposed exist among them. Those who believe that they have a method of isolating a purely internal or a priori component of concepts will want to say that one is capable of inferring a priori from, for instance, "x is a square" to "x is polygonal" because the fact that all squares are polygonal is contained, in one way or other, in his grasp of the concept 'polygonal'. But if one, on the other hand, introspects a state with, say, a dark red quale and exercises the type of recognitional disposition I have been discussing, as a result of which he acquires the concepts 'crimson' and 'red' (as applied to experiences), the fact that whatever feels crimson also feels red cannot be extracted from his grasp of 'red', since there is nothing in his grasp of it over and above his direct acquaintance with the property. Carruthers (2004) shares this understanding of

recognitional concepts, although he prefers to call them purely recognitional, presumably so as to distinguish them from certain ordinary observational concepts the learning of which is assisted by perceptual acquaintance:

A concept is recognitional when it can be applied on the basis of perceptual or quasi-perceptual acquaintance with its instances. And a concept is *purely* recognitional when its possession-conditions...make no appeal to anything other than such acquaintance. A concept is purely recognitional when nothing in the grasp of that concept, as such, requires its user to apply or appeal to any other concept or belief. A purely recognitional concept *of experience* is then a higher-order recognitional concept, which applies to another mental state (*viz.* an experience), and whose possession conditions don't presuppose any other mental-state concepts (not even the concept *experience*) (p. 320, italics original).

Much of the foregoing then points firmly towards ruling out the existence of necessary intralinguistic connections within the introspective domain. It need not of course reinstate the rather stark thesis of atomism across all conceptual domains. Local atomism, as it were, imposed exclusively on the cluster of concepts formed in introspection is as far as one's commitment to recognitional concepts goes.

Some may perhaps embrace this and set out to explain away the intuitive semblance of necessary a priori connections among these concepts by restating them as metaphysically necessary connections. Extensional logic treats extensions of predicates as sets of items which, at a given time, have the properties expressed by the predicates in question. So the inference from "x is crimson" to "x is red" could simply be a matter of the extension of 'crimson' being a sub-set of the extension of 'red' to the effect that the inferential link is accounted for by the fact that it is metaphysically necessary that something which is crimson is also red, for it is undoubtedly its sub-set in every possible world. If the option of establishing a conceptual necessity by an analysis of the purely conceptual components of 'crimson' and 'red' (as applied to experiences) is unavailable because recognitional concepts are stripped of the internal component to be analysed, it nonetheless seems, at a

first glance, that conceptual necessities may in these circumstances be replaced with metaphysical necessities without anything being lost in the process. Fodor, one of the few authors nowadays who thinks it worthwhile to defend semantic atomism, argues in his quite recent book *Concepts* (1998), if I read him correctly, along some such lines.

Concepts which are formed independently of one's other concepts and beliefs and can in principle be in place prior to his learning a language is precisely what the higher-order representationalist needs introspective states aimed at perceptual qualities to embed. In order to appreciate HOR's heavy reliance on the purported autonomy of phenomenal conceptions (i.e. their local atomism), let me summarise the points made in the present section so far. I began by noting that first-order representationalism is ill-equipped to identify the subject of experience, as it does not have the resources to move from "There is a visual representation of something red in me" to "I represent something red". Suppose I only instantiate the first-order functional organisation of a healthy human perceiver, while lacking everything else, so that the former but not the latter description applies to my perceptual situation. Would my perceptual states be qualitative? Would they be experiences? The FOR theorist is bound to answer in the affirmative. But now if we replace the words 'representation' and 'represent' in the two descriptions with 'experience' – a substitution warranted by the claim that experiences are representations – it follows that FOR cannot move from "There is an experience of something red in me" to "I experience something red", which commits its proponents to holding that experiences can occur where nothing is identifiable as their subject, in other words, it commits them to holding that there can be experiences without *experiencers*. This is incoherent, I hope to have shown, and it neglects the fundamental fact that qualia are subjective properties – at least in the sense of 'subject-dependent' properties. Such subject-dependence means that the experiencing self and mental qualities are two sides of the same coin, so to speak, that the self is woven into the experiential fabric, and hence an inquiry into the nature of the concepts that facilitate the self's access to phenomenal properties must figure prominently in an account of why certain representational systems enjoy the privilege of inner life in its full richness while others don't.

But to see that higher-order representationalism stakes its tenability on the claim that the concepts in question are recognitional, we need to be more specific. I contended that,

unlike FOR, HOR merits our approval for doing justice to subject-dependence of phenomenal properties. It distinguishes qualitative sensory representations from non-qualitative ones in terms of availability to introspective or self-reflexive states. This constraint on a state's possession of qualia then banishes the unwelcome implication that experiences can exist in absence of selves which would fill the role of *experiencers*. Causal relations between the self-reflexive faculty (whether thought of as Carruthers' mind-reading module or Dennett's print-out faculty or else) and the various sense modalities, such as those obtaining when an instantiation of a quale elicits the application of a qualitative concept from the higher-order module, are an important building block in the functionalist project of constructing the self from sub-personal material. Recall that Dennett thought it difficult indeed to resist doubts that a system wired up so as to realise a causal network consisting of connections among sub-personal parts would either give us no reason to believe that it is a subject of experience, or would obviously not be a subject of experience. As HOR is a functionalist enterprise seeking to identify the subject of experience, it too is inevitably beset by these doubts, the self escaping its mesh of sub-personal interactions.

Now, it is open to the HOR theorist to postpone addressing these daunting problems directly, and take instead an indirect route via addressing the question as to whether there is anything unique about the concepts a genuine experiencing self forms on being presented with a quale. Given the difficulty of the subjectivity-related issues raised by Dennett, this alternative inquiry should prove more immediately rewarding, as it will at least yield a statement of conditions concerning possession of concepts formed by subjects undergoing experiences. The outcome will then lead to a revised characterisation of the causal relation between the mind-reading module and lower-order perceptual states with a view to building it into the sub-personal causal network from which functionalists seek to construct the experiencing self; so, for instance, if it is a recognitional concept that a qualitative property elicits from the introspective module, then this characterisation will provide some leverage for the HOR theorist in addressing Dennett's worries, since it is far from obvious that a possessor of recognitional concepts does not undergo experiences.

One may retort that to say that something has recognitional concepts is already to presuppose that that something is a subject of experiences. But while this is certainly true, as the characteristics that have been stipulated for them ensure that they invariably pick out phenomenal properties, we mustn't forget that there are other candidates for playing the role of qualitative concepts. Let me explain. Those other candidates are public-language and language-of-thought concepts which differ from recognitional concepts in that their meaning and reference is at least partly determined by their inferential roles (or, if you prefer, by their contingent reference-fixing conditions), and therefore bring with them a wealth of other concepts out of which the premises and conclusions in the meaning-constitutive inferences for a given concept are composed. The higher-order representationalist, however, *must* claim that qualitative concepts are recognitional, with all the characteristics listed in the preceding sub-section and at the beginning of the present one, namely, that they are non-ascriptive, unanalysable, acquired on the basis of brute acquaintance and hence autonomous in relation to any other concepts and beliefs one happens to have. For – and this is the key point – the possession-conditions for, say, public-language higher-order concepts are such that they could in principle be satisfied by a subject who displays perceptual sensitivity to his environment but does not experience anything. Demonstrating that there is a distinct possibility that this is the case with public-language concepts is the central focus of Stephen Leeds' (1993). Bear in mind here that I've been considering HOR as a theory of phenomenal consciousness not as a theory of access consciousness, on which higher-order states realise *phenomenally conscious* awareness rather than afford mere access to a (lower-order) state that is already phenomenally conscious. That the higher-order states accomplishing conscious awareness should contain concepts that can also apply to perceptual states which are not experiences would be something of an embarrassment for its defenders.

Leeds explains that acquisition of public-language introspective concepts – public in the sense that they are used in reports to others and verbal communication in general – is a process that begins by learning, when under a perceptual illusion, to momentarily suspend one's belief that he is under an illusion, and declare instead that the worldly object being perceived does have the property that is non-veridically presented to him. He describes such suspension of belief as 'bracketing of collateral information'.

Reporting one's visual impressions can be taken as a matter of learning to 'bracket' collateral information: instead of reporting the straight stick in water one knows is before one, one learns to report the bent stick one would have reported if one had known nothing about refraction. So if you are to imagine having the ability to report your impressions, what you must imagine is being able to decide certain counterfactuals about your beliefs – e.g. if you had been told that what was before you was a picture, would you have reported it as containing a straight line? I believe that we can in fact imagine this by assimilating it to other cases in which we know – 'just know' – what we would have believed had we not known X (p. 309).

It is in deciding these situations that the question of inner perceptual episodes comes to the fore in one's conceptual scheme. Leeds, unsurprisingly, invokes blindsight as an example of perceiving without experiencing. Since blindsighters are capable of making perceptual discriminations, albeit with much greater difficulty, they too can be subjected to perceptual illusions, and hence it is conceivable, according to him, that they can be trained to bracket collateral information in the above way to enable them to report their 'sensations' in the blind region of their visual field. If, Leeds continues, the only kind of concepts we apply to experiences were such public-language concepts, being affected by blindsight would hardly make the profound cognitive difference we would expect when someone loses a substantial portion of their visual field, since his beliefs would not differ dramatically from ours.

To show in what way we differ from blindsighters, it is not enough to point to *some* difference between their impressions (or *sensa*) and ours: one needs to point to a difference which matters to the character of their mental lives in the way that it matters whether or not we are blindsighters. Blindsighters can identify the missing portions of their visual fields; if they have only recently become blindsighters they know that they are no longer as they were, and long to have

their sight restored: so whether or not one is a blindsighter can make a difference to what one believes and desires. Likewise, of course, we know that we are not blindsighters, and hope we never become like them (p. 310, italics original).

Later in the paper he exploits these considerations in an argument that alongside public-language introspective concepts we, as actually sighted humans, possess concepts which find no expression in public language and which account for the difference between the character of our mental lives and the mental lives of blindsight sufferers. There is little doubt that recognitional concepts are well suited to supply the desired explanation of that difference.

But for our immediate purposes it is sufficient to note that a failure to pinpoint a salient difference between our introspective beliefs and those which blindsighters may be trained to form would render HOR untenable. HOR, examined here exclusively as a theory of qualia, characterises qualia as intentional properties of sensory states available to be targeted by higher-order states. While I viewed the availability condition as a laudable recognition of the fact that experiences are subject-dependent and require experiencing *selves*, qualia would elude the HOR theorist's grasp if the very same introspective states turned out to also target non-qualitative perceptions, that is, if the extensions of the introspective concepts in question included sensory states that are bereft of phenomenal properties. Unless propped up by recognitional concepts, HOR will face the same problem as FOR did in section 1.2 where it was argued, following Carruthers, that its portrayal of qualia could not satisfactorily rule out certain non-qualitative sensory states.

The final reason that advocates of HOR should be eager to hold onto recognitional concepts is that without them they would be compelled to deny raw feels to pre-linguistic children and higher animals. The denial will be forced on them if the mind-reading module which they need in order to identify the subject of experience is formed only upon one's grasp of public-language higher-order concepts. But, on the other hand, if acquisition of qualitative conceptions is achieved purely recognitionally, and thus

possibly prior to and irrespective of the development of other conceptual domains within one's cognition, then they can quite legitimately maintain that the mind-reading module that issues introspective concepts is either hard-wired or develops shortly after birth. Only by showing that self-reflexive states ground recognitional concepts can the defenders of HOR be entitled to credit small children and higher animals with raw feels.

I hope enough has been said in this sub-section to demonstrate that the topic of recognitional concepts in connection with higher-order proposals calls for more than a brief examination, as the nature of higher-order states is not a marginal matter only to be resolved once the vital commitments of such proposals have been made explicit and evaluated. On the contrary HOR does indeed stake its tenability on the existence of recognitional concepts. I have indicated several times that I deny that introspective states feature anything remotely like the concepts described here and in the preceding sub-section; and I'm now going to argue alongside Sellars that qualitative conceptions are composed entirely out of objective and intersubjective elements. These elements, as we will also see, can only be put together by competent public-language speakers. This will result in a view that introspective states are internalised public-language sentences and that there is nothing over and above that which finds perfectly adequate expression in the vocabulary of public language. The consequences of favouring public-language concepts over recognitional ones for HOR and intentional theories of qualia in general will be explored in more detail in 2.3.

2.2.4 The Origin of Qualitative Concepts in Theoretical Discourse: Sellars

Wilfrid Sellars' essay *Empiricism and the Philosophy of Mind* is an attack on classical empiricism and its remnants in Ayer's and Russell's knowledge by acquaintance, both of which doctrines, according to him, conflate non-epistemic episodes of, say, sensing redly with epistemic episodes of knowing that something is red. To say that having a sensation of red either constitutes or entails a non-inferential knowing that something is red is to entangle the idea

(1)...that there are certain inner episodes – e.g., sensations of red or of C# which can occur to human beings (and brutes) without any prior process of learning or concept formation; and without which it would in some sense be impossible to see, for example, that the facing surface of a physical object is red and triangular, or hear that a certain physical sound is C#,

with the idea

(2)...that there are certain inner episodes which are the noninferential knowings that certain items are, for example, red or C#; and that these episodes are the necessary conditions of empirical knowledge as providing evidence for all other empirical propositions (Sellars, 1997, p.21).

It is, in other words, to assimilate phenomenal consciousness to cognitive or conceptual consciousness. The former doesn't entail (let alone constitute) the latter because conceptual awareness is classificatory – as Richard Rorty explains in the commentary on Sellars that appears in his (1979), according to Sellars' Wittgensteinian view, knowing what X is comes down to knowing what kind of thing it is. We do not learn the concept of X simply by 'noticing' X when in perceptual contact with its instance, which would be the case if (1) entailed (2). To notice X is to notice it under a certain description, which description presupposes understanding that X belongs to a certain class of things the concept or concepts of which we must already possess. So having the concept of X is being able to inferentially relate that concept to the numerous other concepts contained in propositions used by speakers as premises in justifying their beliefs about X.

The fact that cognitive consciousness is essentially a classification of things under kinds, requiring that a speaker be able to link his newly acquired concepts with other concepts in the practices of reasoning and justification is, in the broadest outline, Sellars' reason for endorsing a version of inferential role semantics, one where the appropriate rules of inference that govern these practices are material rather than formal. Rorty

warns, however, of interpreting Sellars as making the case that pre-linguistic children do not feel pain until they learn the concept of pain. They feel pains and react to them because they *have* them. What they learn later in their lives is to know what kind of thing pains are. Small children of course respond behaviourally to scalds, tingles, itches, etc., but it is not until they have a grasp of the relevant vocabulary that categorises these occurrences, and participate in reasoning about them with other members of their linguistic community that we can say that they know what they are.

Sellars does not deny that there are non-inferential beliefs, if ‘non-inferential’ is understood in the epistemologically benign sense of ‘observationally elicited’. But he would stress that non-inferential uses of concepts (contained in observational beliefs) in this sense presuppose that one is already able to use them inferences. What he rejects is that there are non-inferential beliefs in the sense that one can arrive at them solely by having sensations, even though he has no other concepts and beliefs to which his observational concept is inferentially linked. Indeed, on Rorty’s reading, Sellars would see no reason not to grant that a congenitally blind person may learn to use ‘red’ inferentially to the effect that he will know, for instance, that red is a colour, that it differs from blue, green, etc., and hence know what it is in the sense of knowing what *kind* of thing redness is, despite not being able to apply it non-inferentially in response to perceptually presented states of affairs. Sellars could grant all of this while maintaining that sighted pre-linguistic infants do not know what redness is.

The suggestion that small children must know what redness is just by being awake amounts to a confusion, nurtured by classical empiricism, of perceptual sensitivity with genuine conceptualisation that results in a view of perceptual concepts as labels that attach to aspects of sensory contents once the channels of perceptual inflow are opened without their prior employment in reasoning (i.e. inferences). If conceptualisation was separable from reasoning, justification and inferring, if non-inferential applications of concepts were not mere offshoots of inferential ones, there would be nothing to define us as rational and sapient beings, for non-inferentially formed ‘labels’ (‘non-inferentially’ in the sense that Sellars rejects) would scarcely differ from the outputs of photosensitive devices or the noises that parrots can be conditioned to make. The following sums up the position:

The essential point is that in characterizing an episode or a state as that of *knowing*, we are not giving an empirical description of that episode or state; we are placing it in the logical space of reasons, of justifying and being able to justify what one says (Sellars, 1997, p.76, italics original).

Although so far I have only talked about concepts referring to sensory contents and there has as yet been no mention of introspective concepts referring to sensory states and their properties, it is clear that if the thought that non-inferential uses presuppose inferential ones is carried over to higher-order concepts, it will turn the story of recognitional concepts on its head. A recognitional concept is installed in one's cognition when a causal relation with a phenomenal property is established and does not owe its epistemic status and reference to having a role in reasoning, which secures that only sighted persons are in a position to satisfy its possession-conditions. Again, this is precisely the story of non-inferential concept acquisition that Sellars would dismiss as an attempt to empirically describe an occurrence that derives its epistemic status from its existence within the logical space of reasons.

These opening paragraphs are to serve as an outline of Sellars' general approach to so-called 'ground-level' beliefs about aspects of sensory contents taken by classical empiricists to be non-inferentially individuated, where the supposed non-inferential individuation was viewed by those philosophers as a means of delineating a privileged class of propositions that are the source of justification for all other empirical propositions without themselves standing in need of justification. The empiricist project of laying a foundation for the entire body of empirical knowledge is a manifestation of the Myth of the Given which Sellars dismantles in his work.

The distinction between inferential and non-inferential uses of concepts will, however, continue to bear heavily on his treatment of introspective concepts, except that for the introspective domain it is drawn so as to have the force of the distinction between theoretical and observational uses – theoretical concepts are about entities that are unobservable at a certain stage in the development of a given field of scientific study and

therefore occur in statements that are only applied as inferences from statements of observable fact. But such entities can always become non-inferentially (in the sense of ‘observationally’) reportable if, for example, someone with a strong background in physics gradually (and unwittingly perhaps) abandons the vocabulary of common objects in favour of the vocabulary of a confirmed physical theory and instead of reporting, say, a bolt of lightning in the sky he is now inclined to proclaim that he sees an electric discharge in the sky. In Sellars’ view, the report he is now inclined to make has an equal claim to be non-inferential or ‘ground-level’. It does not differ in status from the one he has abandoned insofar as ‘electric discharge’ has come to play the role in his conceptual scheme that ‘bolt of lightning’ does in the conceptual scheme of lay people. That introspective concepts undergo similar shifts from theoretical to observational discourse is Sellars’ most notable contribution to theorising about introspection. Further clarification of the point is due below.

I shall start with an argument that gets underway in section 10 of his (1997) and is closed in section 22 against the idea – common to cartesians and empiricists – that sensory contents or appearances are given in that they are objects of immediate awareness on which knowledge of physical objects in space and time rests. This particular argument, unlike the outline of his general approach just above, does not depend on our prior acceptance of Sellars’ normativist (as opposed to computationalist) version of inferential role semantics. Its primary purpose is not a defence of inferential role semantics, although a rejection of conceptual nativism and atomism naturally flows from it, but to bring to surface the first dimension of intersubjectivity built into concepts of experiential inner episodes, and introduce the terminology from which, as we will appreciate later, full-fledged introspective concepts eventually evolve. I will then be in a position to focus on the key sections of his essay, those where he rejects the idea that experiential inner episodes are given in that they are objects of immediate awareness, that one non-inferentially knows that they occur just by having sensations, *something which would be the case if Loar’s recognitional concepts were possible*. At that stage the second dimension of intersubjectivity built into introspective concepts will emerge.

Philosophers in the cartesian and empiricist traditions alike took it that beliefs about physical objects in space and time are inferences from direct and infallible knowledge of

sensory contents or so-called appearances. Knowledge of facts such as X appears red is, according to them, a matter of the most fundamental and pre-linguistic awareness from which, if the circumstances of perception are regarded as standard, a perceiver concludes that X also is red. By making the inference a perceiver leaves an ideal world of pure appearances known to him intimately and without mediation, and enters a world of physical objects in which he is inevitably prone to error. Hence the notion 'is red' is thought to be reducible to 'looks red'. Sellars' considers the following statement:

(D) x is red \equiv x looks red to standard observers in standard conditions

Those who are the target of his criticism would see the right-hand side as the definition of 'is red', that is, as the definition of physical redness in terms of 'looks red'. In response to this misconception, he first distinguishes three kinds of states:

- (a) Seeing that x , over there, is red
- (b) Its looking to one that x , over there, is red
- (c) Its looking to one as though there were a red object over there

Although he calls these states 'experiences', which slightly obscures his exposition, it is clear enough that the term is not used in his text with the meaning it typically has in contemporary literature. What he means here, if I read him correctly, is something we may call a 'total experiential state', one that comprises a perceptual belief occasioned by what would nowadays be referred to as 'experience' or 'perception' on the one hand, and an impression or immediate experience on the other, which are Sellars' equivalents of experiences and perceptions as they are currently understood. It comprises something belonging to the epistemic realm as well as something belonging to the causal realm. We could certainly imagine a little more fortunate choice of terms, but we ought to bear in mind that his problem is whether the acquisition of concepts pertaining to sensory contents precedes the acquisition of concepts pertaining to physical objects, so including epistemic (and hence *conceptual*) episodes under the term 'experience' perhaps helps bring out the point of disagreement between him and the approach he criticises.

Now, on Sellars' diagnosis, the above misconception results from an oversight of two facts. Firstly, to characterise an experience (in his sense) as that of seeing that something is the case is to ascribe to it (or to its epistemic ingredient) a certain assertion or propositional claim, namely, that *x* is red. Elsewhere he calls it 'propositional content'. Secondly, and more importantly, it is also to say that its subject *endorses* the claim. Here he refers to Ryle and likens this notion of endorsement to his emphasis that 'see' is an achievement verb. Ryle's insight, according to him, is that in describing an experience as an instance of seeing that something is so and so we "apply the semantical concept of truth to that experience" (1997, p. 40). Thus one is compelled to judge "I see that *x* is red" instead of the mere "*x* is red" when "certain considerations have operated to raise, so to speak in a higher court, the question 'to endorse or not to endorse'" the experience's propositional content '*x* is red' (p.41). This suggests to Sellars that "*x* looks red to me now" belongs on the same level as "I see that *x* is red", for, again, one is disposed to judge "*x* looks red to me now" only when the question 'to endorse or not to endorse' has arisen, save that now he withdraws his endorsement of that very same propositional claim (i.e. '*x* is red') that his experience involves. We are disposed to judge "*x* looks red to me now" rather than "I see that *x* is red" when an object is viewed in, say, unusual artificial light and we suspect that it may not appear what it is. Nonetheless, in making such a report, it is '*x* is red', where the redness is physical redness, not one ascribed to sensory contents, that we recognise or even initially assent to, but subsequently hold back from endorsing.

So experiences (a), (b) and (c) share at least those parts of their respective propositional contents which may be expressed as '*x* is red' and 'there is an *x* over there', the difference among them consisting in which parts of the propositions, if any, we are prepared to endorse. In (a), we commit ourselves to both '*x* is red' and 'there is an *x* over there', in (b) the former falls outside the scope of the endorsement, and neither is endorsed in (c).

We can now understand where this reasoning parts ways with the classical positions scrutinised by Sellars. In fact, if it is sound, it rebuts their most basic presuppositions. Since '*x* is so and so' is embedded in 'looks so and so', it implies that we do not conceptualise aspects of sensory contents until we can conceptualise properties of

physical objects. In classical accounts, however, undue importance was attached to the thought that because we seem to have infallible (or incorrigible) knowledge of sensory contents, that we cannot be wrong about (or be corrected on) how things appear to us, concepts pertaining to sensory contents or appearances are formed solely by virtue of having sensations and must lie at the base of our entire body of empirical concepts. The reasoning is also invoked in *Phenomenalism* (Sellars, 1963) to argue against a cluster of positions (including the sense-datum theory) which are in agreement with the classical conceptions about there being a more basic form of visual knowledge than seeing *physical objects*, but conceive that level of basic knowledge as consisting of sensings of two-dimensional colour expanses floating before an inner sense. Sellars now returns to (D), his original formulation of the problem,

(D) x is red \equiv x looks red to standard observers in standard conditions

and concludes that it is a necessary truth

...*not* because the right-hand side is the definition of “ x is red,” but because “standard conditions” means conditions in which things look what they are. And, of course, *which* conditions are standard for a given mode of perception is, at the common-sense level, specified by a list of conditions which exhibit the vagueness and open texture characteristic of ordinary discourse (1997, p. 43-44, italics original).

The discussion reveals the first dimension of intersubjectivity upon which, as we will appreciate shortly, introspective conceptions are built. For its gist is that the framework of sensory contents rests on the undeniably *intersubjective* framework of physical objects in space and time.²² But before I can explain precisely how this particular intersubjective element is involved in introspective conceptions as well as how its involvement

²² In fact Sellars goes as far as to suggest that the inverted spectrum hypothesis too relies on the mistaken assumption that we form beliefs about appearances or sensory contents independently of and prior to grasping concepts of the properties of worldly objects. The possibility of inverted spectra would not be raised if the role of intersubjectivity in talk of appearances was properly taken account of.

undermines the claims that introspective concepts are recognitional, we must pick up the line of argument that constitutes Sellars' denial of the givens of sense impressions as objects of immediate awareness. Its final step is completed when he shows how – once I possess the concepts of *seeing* that x, over there, is F, *its looking* to me that x, over there, is F, and *its looking* to me as though there were x that is F over there – experiential inner episodes or impressions are extracted within my conceptual scheme as the common components of such families of states as the (a), (b) and (c) above.²³ Sellars also refers to this common ingredient as 'descriptive content' shared by the three situations alongside their propositional content (i.e. 'x is F' and 'there is an x over there') which we either endorse or refuse to endorse (or parts thereof). It is the common causal antecedent of the corresponding thoughts "I see that x, over there, is F", "X, over there, looks F" and "It looks as though there were an x that is F over there".

So the latter sections of his (1997) are an attempt to trace, with the aid of a fictional story about our Rylean ancestors, a process of intellectual development that begins with the formation of concepts referring to the perceptible properties of physical objects and culminates in the formation of concepts whereby we recognise each other and ourselves as creatures who think, have feelings, sensations etc., and who can 'observe' and be directly aware of the items in their stream of consciousness. For Sellars, the need for clarifying the status of impressions by concentrating on their development arises from the realisation that

...once we give up the idea that we begin our sojourn in this world with any – even a vague, fragmentary, and indiscriminating – awareness of the logical space of particulars, kinds, facts, and resemblances, and recognize that even such "simple" concepts as those of colours are the fruit of a long process of publicly reinforced responses to public objects (including verbal performances) in public situations, we may well be puzzled as to how, even if there are such things as impressions or sensations, we could come to know that there are, and what sort of

²³ According to Sellars, having the concept of seeing that something is so and so together with the concepts of qualitative and existential lookings alone would only enable a perceiver to refer to his experiential inner episodes indirectly, by saying that if the common propositional content of, say, (a), (b), and (c) was true, they would all be instances of seeing that x, over there, is red. A further enrichment is necessary in order for someone to be able to focus introspectively on his sense impressions.

thing they are. *For we now recognise that instead of coming to have a concept of something because we have noticed that sort of thing, to have the ability to notice a sort of thing is already to have the concepts of that sort of thing, and cannot account for it* (p. 87, italics original).

The first part of the story that recreates sense impressions from a base made up of the ground-level observational concepts of public objects concerns thoughts. Thoughts are examined at the outset because, as I already hinted above, impressions are initially conceptualised, or first ‘noticed’, as it were, by someone who has no previous theory of mind (whether folk or other) with respect to perceptual situations, as the common causal antecedent of the *thought* “I see that something is so and so”, and the corresponding *thoughts* regarding qualitative and existential lookings. Furthermore, thoughts too are treated as theoretical posits which became observable and hence follow a basic developmental pattern that is repeated in the case of experiential inner episodes.

Sellars introduces an imaginary community of our ancestors whose language is Rylean in that it is confined to concepts of public objects and properties. In addition to these, it contains logical operators and whatever other grammatical tools are necessary for construction of sentences, but is devoid of mentalistic concepts to the effect that its members lack a theory of mind. We can imagine that the situation being described here is loosely analogous to that in which infants find themselves early in their intellectual life when they conceptualise items in their surroundings but respond to scalds, thirst or sleep deprivation only in an instinctive and pre-conceptual way. When the members of the community try to explain their actions and verbal responses as well as those of others, they do so not in the familiar mentalistic vocabulary that we frequently make use of (for example, “I ran because I *thought* I was being chased”), but in the vocabulary of behaviouristic psychology. Now, the remainder of the essay is an answer to the question as to what additional practices they must adopt to be able to view the actions and verbal performances of others as well as their own as being brought about by inner occurrences known to us as thoughts, feelings, sensations, etc.

As for thoughts, the answer has two parts. Firstly, it is the application of semantic categories to their utterances, the practice of their semantic evaluation, of clarifying their

meaning when ambiguities obscure it, of ascertaining their truth or falsity, etc. Semantic categories are indispensable because thoughts (which of course they do not yet conceptualise) inherit their 'aboutness' from the semantic features of overt speech. Linguistic meaning is explanatorily more basic than intentionality. We cannot think what does not find expression in our public-language vocabulary. This is clearly a non-mentalistic approach to the intentionality of non-phenomenal mental states as contrasted with, for instance, the functionalist movement initiated in the early work of Putnam (1960, 1967). The latter is a mentalistic strategy since it construes linguistic performances as expressions of thoughts that are individuated (or at least their narrow component) by their computational role within a causal network, not by their correlation with utterances which, on this position, may or may not express them adequately. The causal network must be complex enough to confer a sufficiently determinate content on various thought-states, but, crucially, it can be in place regardless of whether a subject has learnt to speak. Many have urged with this in mind that concerns about meaning should be taken over by theories of the content of thoughts or propositional attitudes because meaning is parasitic on intentionality.

On the non-mentalistic picture that is beginning to emerge for Sellars, linguistic acts are not expressions of independently formed thoughts, rather, thoughts are *internalisations of public-language sentences*. The internalisation is prompted in an individual in the following way: when an infant (at the initial Rylean stage) acquires his first language, he does what we, having the concept 'thought' already, would call 'thinking-out-loud'. Later he learns to resist the compulsion to speak so he can think without thinking-out-loud. The idea that thoughts are the outcome of suppressing one's compulsion to speak is spelt out in his (1969).

Let us suppose then that the members of the originally Rylean community are now capable of accomplishing thought processes and that their intelligent activity is not restricted to sequences of thinkings-out-loud. Although they have become creatures who think in precisely the same way as we do, this internalisation alone does not make their thoughts introspectible. The key breakthrough in their ascent from behaviouristic psychology to the ordinary mentalistic discourse comes with the addition of the practice of postulating unobserved theoretical entities (e.g. molecules) in order to explain the

behaviour of such observable phenomena as temperature or gas pressure. Inner occurrences are introduced into their conceptual scheme when they realise that their fellows behave intelligently even when their actions are not accompanied by overt speech, and hypothesise on that basis that it is episodes of ‘inner speech’ that lie at the beginning of the (downstream) causal chain which steers the course of their action. That they are first understood by analogy with inner speech is to give sense to the earlier suggestion that thoughts owe their intentionality to the semantic properties of verbal performances. Just as before our fictional ancestors started to engage in assessing the truth and clarifying the meaning of each other’s (observable) linguistic acts, and came to think of them as being about such and such, they now think of these (unobserved) theoretical postulates as being about such and such. Thus the analogy is merely to highlight the fact that the semantic features of overt speech are carried over to these inner episodes, it is not to imply that these episodes are “the wagging of a hidden tongue, nor are any sounds produced by this “inner speech”” (1997, p. 104).

Such newly posited entities are not yet thoughts as we know them, for they are in intelligent creatures as molecules are in gases. Neither are they in them as ghosts are in machines. Like sub-atomic particles, for example, they are unobservable in that they can only know about them by means of an inference from observable behaviour, whether their own or that of others. Their knowledge of them is always a conclusion of an inference. And like all other theoretical, as opposed to empirical concepts, they lack non-inferential or observational uses. Thinking that they have privileged and infallible access to their inner episodes would be deemed, even from their own perspective, as absurd as thinking that we have privileged and infallible access to electromagnetic waves.

At this penultimate stage our fictitious ancestors frequently appeal to episodes of ‘inner speech’ in interpreting the behaviour of others. And they finally become *introspectible thoughts* once they have learnt to non-inferentially (hence observationally) report their presence in themselves. What was until now only known inferentially – if, for example, a person wishing to quench his thirst by reaching out for a glass of water said “There is a glass of water in front of me”, his knowledge of the corresponding ‘inner utterance’ that had caused him to act as he did was an inference from his overt behaviour – is now non-inferentially reportable: “I’m reaching out for a glass of water because I’m

in a state of *thinking*, of which I was directly aware prior to any action or verbalisation, that there is a glass of water in front of me”.

But more needs to be said about the final step. For Sellars, it is but a mundane case of inferentially known theoretical entities breaking into observational discourse. The gap between theoretical entities and observable entities is purely methodological and carries no ontological import. One useful example is due to Robert Brandom in his (Sellars 1997) commentary:

Thus when first postulated to explain perturbations in the orbit of Neptune, Pluto was a purely theoretical object; the only claims we could make about it were the conclusions of inferences. But the development of more powerful telescopes eventually made it accessible to observation, and so a subject of non-inferential reports. Pluto did not undergo an ontological change; all that changed was its relation to us (p. 164)

There is no rigid boundary between what is observed and what is inferred. Moreover, to get to the heart of the matter, shifts from theoretical to observational discourse need not be facilitated by improving our powers of observation and the ability to report our thoughts is certainly not the result of a fine-tuning of an ‘optical’ instrument in the mind. Indeed, these shifts are very often a matter of conclusions of inferences gaining a reporting role. For there are circumstances in which, say, a chemist is disposed to state that he is looking at sodium chloride without thinking of the substance as the common flavour-enhancing ingredient in food, even though outside his laboratory he would normally infer that statement from his seeing that there is salt on the table. What is applied observationally in one set of circumstances may be applied inferentially in another. The chemist’s non-inferential use of ‘sodium chloride’ does not mean that his eyesight is so superior as to enable him to see molecules of sodium chloride in salt crystals, it only means that the specialist vocabulary of chemical elements and compounds has now, at least in his conceptual hierarchy, established itself at the observational level alongside the common-sense vocabulary (e.g. ‘salt’, ‘water’, etc.) used by lay persons.

Once we rid ourselves of the foundationalist misconception that knowledge of appearances is a supreme example of observational knowledge in that statements about sensory contents do not stand in need of justification but are the justificatory roots of all other statements of empirical fact, nothing is observational in an absolute sense and it is therefore reasonable to demand a justification of any empirical claim. Take, for instance, colour predicates applied to physical objects. Think of a situation in which a perceiver is challenged to justify his belief that that the ripe tomato in front of him is red, and does so by saying that it is being viewed in standard conditions and that in standard conditions he is unlikely to be mistaken. Likewise, as we have seen, nothing which really exists is, in this context, concealed from observation if ‘observable’ simply amounts to ‘non-inferentially reportable’.²⁴

Indeed, in Sellars’ sense, one who mastered reliable differential responsive dispositions noninferentially to apply normative vocabulary would be directly observing normative facts. It is in this sense that we might be said to be able to *hear*, not just the noises someone else makes, but their *words*, and indeed, *what they are saying* – their *meanings* (Sellars 1997, p.165, italics original).

Now that our ancestors introspect their thoughts and conceive them as occurrent or episodic rather than dispositional, the philosophically-minded members of the community will perhaps want to add that they are also private. But it is a sense of ‘private’ that has been robbed of its traditional force. Thoughts are originally introduced as common-sense theoretical posits in order to identify the true source of intelligent behaviour in situations when it is unaccompanied by overt speech. So in its nascence the concept of a thought is as intersubjective as the concept of an electron, which intersubjectivity remains embedded in the full-fledged introspective concept that is formed later. The privacy of inner episodes, from a subject’s point of view, consists in little more than the fact that, unlike other familiar examples of theoretical entities, they are his own states. In other words, the respect in which the sentence “S is thinking p” stands out in comparison with

²⁴ I owe this extended interpretation of Sellars’ view of the distinction between theoretical and observational discourse to Robert Brandom’s study guide in the edition of *Empiricism and The Philosophy of Mind* that I consult here.

sentences in the language of the kinetic theory of gases is that it is logically true of it that “whereas *anybody* can use it to state a fact, only *one* person, namely S himself, can use it to make a report (Sellars, 1997, p. 87, italics original).

Another point that considerably weakens the appeal of the traditional understanding of privacy is that one’s employment of mentalistic vocabulary in self-description presupposes the ability to interpret in mentalistic terms the behaviour of others. Hence conceptual responses to inner episodes, as we saw above, can be said to result from projections from the case of others to our own case, not the other way round. Standing in opposition to this is the construal of privacy that is sustained by the supposition that thoughts, feelings and sensations are simply laid out before the mind’s eye, that we notice them in a primitive and pre-linguistic way, and that our knowledge of other minds is a projection of concepts that are the products of such acts of private ostension or noticing. These last remarks will be repeated in the discussion of sense impressions and are of great importance as they mark one of the key points where Sellars’ approach clashes with Loar’s contention that introspective concepts are recognitional.

Having recreated concepts referring to thoughts from intersubjective material (i.e. overt speech or ‘thinkings-out-loud’, theoretical entities and the observable behaviour of others) while dispensing with conceptual nativism and private ostension, he now vows to do the same for concepts referring to sense impressions.

He first returns to his earlier analysis of qualitative and existential lookings in terms of perceptible properties of physical objects and scope of endorsement. The upshot was that the possession of concepts referring to public objects and properties is a necessary condition of our grasp of concepts to do with appearances or sensory contents. Families of perceptually elicited thoughts such as (a) seeing that x, over there, is red, (b) its looking to one that x, over there, is red, and (c) its looking to one as though there were a red object over there, contain the same proposition, namely, “X is red” (where ‘red’ refers to a public property) which is endorsed in (a) but fails to be endorsed in (b) and (c). Since our ancestors now have a common-sense theory of mind concerning thoughts, they can also introspect perceptual thoughts of those three types. Sense impressions are discovered in their attempt to explain why there appears to be a red object over there in each of the perceptual situations that lead them to think (a), (b), and (c) respectively

despite the apparent variations in external circumstances. They hypothesise that there is an inner episode which differs from thoughts in that it belongs to the causal rather than epistemic realm, and is the common causal antecedent of (a), (b) and (c).

Given that they, like thoughts, are states rather than particulars, they are not literally coloured. What is more, we must not forget that this is a reconstruction of *ordinary* mentalistic terms and in ordinary talk we only speak of worldly objects as being literally coloured, although impressions resemble and differ in a system of ways which is “structurally similar to the ways in which the colours and shapes of visible objects resemble and differ” (Sellars, 1997, p. 112). Impressions are coloured in ways which are analogous to the colours of physical objects. Sketchy though it seems, this characterisation does no violence to the vagueness of the ordinary-talk terms we use to communicate our feelings and sensations to others. We use them with much authority without systematically probing beneath the surface of the entities in question in the way philosophers and psychologists do. All that an examination of the common-sense framework (what Sellars also calls ‘the Manifest Image’) reveals is that we do not conceptualise sense impressions as particulars, let alone as particulars that are literally coloured, since within that framework only physical objects are thought of as being literally coloured.

To sum up, the purpose of positing sense impressions in ordinary talk is to throw light on situations where, for instance, there appears to be a red object over there even though the perceiver has solid grounds for believing either that the object being viewed is not red or that there is no object at all occupying that particular position in space. We do not simply notice sense impressions in ourselves but initially appeal to them as physicists do to sub-atomic particles in rationalising other persons’ verbal or behavioural responses to misperceived or hallucinated states of affairs. In short, as in the case of thoughts, we proceed from third-person ascriptions to first-person ascriptions. At the beginning first-person ascriptions are conclusions of inferences and it is only after further mutual conditioning and inculcation has taken place that we come to be directly aware of them, just as a chemist with a sufficient amount of training develops a disposition to non-inferentially state that he is observing sodium chloride being dissolved in H₂O. So the additional conditioning and inculcation gradually compresses, as it were, the however

many inferential steps are required to reach a conclusion about a theoretical object up until the point when it is reported observationally. Nevertheless, in accordance with inferential role semantics, non-inferential uses of concepts that previously denoted unobserved entities continue to presuppose and be individuated by inferential ones.

This concludes Sellars' reconstruction of introspective conceptions of experiential inner episodes from propositions about observable physical entities of the form "X is F" on the one hand, and certain propositions about unobserved theoretical entities (later appealed to in self-description) on the other. The former come into play by being embedded in perceptual thoughts about qualitative and existential lookings (but fall outside the scope of endorsement), the awareness of which thoughts prompts the positing of experiential inner episodes. Both are undeniably intersubjective and public – 'theoretical' hardly implies 'private'. Laying bare the dependence of introspection on the conceptual domain pertaining to public objects and properties – whether observed or theoretical – has the merit of dispelling the concerns voiced by some wittgensteinians that we can have no knowledge of inner occurrences since, being private, they systematically elude public discourse. These complaints turn out to be an overreaction. It also casts doubts over claims that inner occurrences are ineffable; again, 'theoretical' hardly implies 'ineffable'. The privacy of experiential episodes consists in little more than the fact that it is logically true of sentences such as "S has a toothache" that whereas anybody can use it to state a fact, only one person, namely S himself, can use it to make a report. From the subject's perspective, experiences differ from other examples of theoretical objects in that they are states of his own.

Now that their progression from the Rylean stage to ordinary mentalistic talk is complete, some of our philosophically-minded ancestors, having lost sight of how they have come to see themselves as creatures who have sensations and feelings, will be tempted to say that their higher-order concepts are recognitional, that they formed them on the basis of brute acquaintance with their experiences, that their mode of presentation is exhausted by the causal relations with their referents, as nothing is closer to the mind than the mind itself. Others may press the case by pointing out, together with Levine, that the concepts in question have non-ascriptive modes of presentation, or, together with Loar, that they do not refer via a contingent reference-fixer. There are, they would

continue, no *constitutive* inferential connections with other concepts to which they at least partly owe their meaningfulness and such autonomy is perhaps the sole preserve of introspective conceptions given that nothing is more immediately and intimately known to the mind than what it itself comprises. It is for this reason that I have called this position 'local atomism'.

Surely, we may not agree with every detail of Sellars' story, but it is sufficient if we acknowledge, as we should, that there are no experiential states to notice prior to their serving an explanatory purpose and our recognition of that purpose. The alternative we must distance ourselves from

...makes the mistake of supposing that the logical space of the concept simply transfers itself from the objects of direct perception to the intellectual order, or better, is transferred by the mind as Jack Horner transferred the plum (Sellars, 1963, p.334).

To further sharpen the contrast between theory-laden introspective concepts and recognitional introspective concepts, note the following remarks by Loar:

In some cases, a recognitional disposition or ability is a disposition or ability to apply an *independent* term or concept, one whose initial mastery does not involve a specific recognitional ability. For example, a person might come to understand 'porcelain' from a technical description and only later learn visually, tactually and aurally to identify instances. That is not the phenomenon I mean (Loar, 1990, p. 88, italics original).

Setting aside, of course, the fact that introspection is not a matter of *perceptual* identification on anyone's view, it is evident that it is some such phenomenon that Sellars means. This is not to deny that, for a certain class of concepts, acquisition requires perceptual acquaintance or recognition. For instance, one is not fully in possession of the

concept 'red' until he reliably identifies the property in his surroundings. However, it can not be emphasised enough that the kind of recognition in question is already an exercise of concepts to the effect that in order for someone to succeed in visually identifying redness he must know at least that it is a visual property, that it is a colour, that it differs from blue, yellow, green, etc. What is the source of contention – and what Sellars would deride as epistemological fiction – is the *purely* recognitional capacity appealed to by Loar and others, the exercise of which does not presuppose conceptual classification, thereby turning out concepts whose mode of presentation reduces to the causal link with their referents.

Suppose, for the sake of exposition, that it is possible to flout Sellars' justificational holism about concepts of worldly objects and prove beyond doubt that their logical spaces do after all transfer themselves from objects of direct perception to the intellectual order. Nevertheless, it would remain utterly mysterious how anything of this nature could also hold for introspective concepts – however hard I try to focus inwards, *I fail to be acquainted with anything in a way that is even remotely analogous to 'outwardly directed' perception*. And to postulate a dedicated organ of inner sense would be to slip further into the obscure. Therefore the current scenario still leaves an explanatory void to be filled by a theory which reconstructs higher-order awareness rather than assumes its givenness.

Let us now return briefly to the argumentation which prompted the examination of recognitional concepts. Drawing heavily on Carruthers' presentation of evidence to do with non-qualitative route of perceptual processing in healthy humans (which is thought to be preserved in blindsight sufferers and explains their remarkable ability to make out certain features of distal stimuli), it was argued against first-order representationalism that it is excessively inclusive in that, given its statement of conditions for phenomenality, these representations too qualify as qualia-bearing. What is more, as I reiterated in several places, organisms which utilise representations of distal stimuli without undergoing experiences may arguably be a common occurrence in nature. Recall that in 2.1 Tye's poisedness was deemed ineffective in narrowing down the class of perceptions that satisfy the condition to qualitative ones. I then agreed with Carruthers that, in view of these difficulties, it is experiential subjectivity (as distinct from worldly

subjectivity which non-qualitative and qualitative perceptions may exhibit in equal measure, provided the former too represent distal rather than proximal stimuli) that suggests itself as the criterion that adequately marks out the set of qualia-bearing representations. That criterion in fact comes down to introspectability. Thus higher-order representationalism was endorsed as a substantially more robust position.

Introspectability, or availability to self-directed states, took centre stage again in a quite independent consideration in 2.2.1. It revealed more deficiencies in FOR that worked in favour of HOR. In a nutshell, phenomenality and the self, what Dennett (1978) calls the exempt agent, are inextricable. That the self is, as it were, immersed in ‘mental paint’ is among the most fundamental facts about phenomenality and it applies with just the same force to non-conscious (to mean unattended) qualitative representations – hence the looser requirement that qualitative representations be *introspectable*, not *introspected*. Being first-order, FOR’s statement of necessary and sufficient conditions for phenomenality makes reference exclusively to relations at the lower-order level (plus external inputs). By severing the inextricable connection, it puts the self in the position of an observer rather than a subject, according qualia, I argued, the kind of independent status that only naïve realists would be content with. The result is an inconsistent jumble of ideas ranging from relationalism to naïve realism.

While the self will prove to be a thorny matter for anyone whose aim is a fruitful enquiry into how it came into existence in the natural order, what seems certain is that any such enquiry cannot dispense with a system of metastates. It is here that we noticed that HOR is tethered to recognitional concepts. Because the meaning of recognitional concepts reduces to the causal relation established in the ‘act’ of pure recognition, it is out of the question that blindsighters could deploy them in response to stimulation in their blind region. Unfortunately the problem is that to think that these are genuine and full-fledged concepts, i.e. items existing in the epistemic realm, is to repeat the mistakes of classical empiricism which largely due to Sellars’ extensive criticism and the virtues in his alternative treatment cannot be ignored.

Before I can reach the conclusion that has been the aim of my present discussion, namely, that

(iii) to have a higher-order thought about one's perceptual state is to token an internalised public-language sentence whose meaning is holistic and normative,

and hence that there are no concepts which distinctively apply to phenomenal properties, I should dedicate a few paragraphs to a response which merits interest for not resting on the idea that we acquire concepts upon noticing objects and properties or that there is a concept-forming act of pure recognition, while leaving sufficient room for phenomenal concepts. The mention of normativity in (iii) also obliges me to give some explanation of why it is inevitable that if the introspective concepts in question are internalised public-language concepts, they will be enriched with a normative dimension. The next subsection will be a convenient place to do so as seeing why the response is more underwhelming than it first appears motivates a straightforward case for normativity.

2.2.5 (Innate) Mentalese Concepts?

The response I now wish to assess from a Sellarsian standpoint is outlined in the article by Stephen Leeds' (1993) that helped develop the topic under discussion (see 2.2.3). Leeds is wholly convinced by what he calls 'qualia intuitions' and believes that there are phenomenal concepts. The idea that there may be beings who display behavioural sensitivity to visual stimulation without undergoing experiences is comprehensible to us all. Yet we know we are not such beings, or blindsighters, for that matter. Thus it must be perceptual beliefs containing phenomenal concepts that inform our knowledge that we happen to be neither of the two.

His proposal is that phenomenal concepts are a special class of Mentalese concepts – a possibility which Sellars is more or less silent about but would not necessarily want to resist. For Leeds, as I'm about to explain, thinks that Mentalese phenomenal concepts construed in a certain specific way are amenable to being at least partly combined with justificational holism. Nothing along these lines would have been feasible in the case of recognitional concepts – it is out of the question that one could argue for the existence recognitional concepts and not run afoul of the basic principles of justificational holism.

He also makes it clear that it is not his ambition to use phenomenal concepts as a material for laying a foundation for the rest of our knowledge.

The core steps in the argument that if there are phenomenal concepts, they must be lexical items of Mentalese appear in the final part of his text where he first explores a number of ways a technical spoken-language word expressing a phenomenal concept such as 'looks red' or 'phenomenally red' can come to have a meaning (he acknowledges that phenomenal concepts have no expression in ordinary language, hence the introduction of a technical term). Given his earlier argument for the existence of phenomenal concepts, he now simply assumes that there is an underlying phenomenal concept which lends 'looks red' its semantics. I shall do the same for the sake of my own argument. Agreeing with Sellars that we do not notice and name qualitative properties without prior conceptualisation, he would deny that the concept in question attaches to its referent in the same way as purely recognitional concepts do. Nor, needless to say, can it be the kind of intersubjective concept reconstructed by Sellars. According to Sellars, phenomenal states and their properties are conceptualised as unobserved theoretical postulates in order to explain why, for instance, objects placed in front of a mirror also appear to be behind it, or why some surfaces appear to change colour under artificial light. We later develop a disposition to report them directly in the same way as we may be trained to report rapid molecular motion instead of heat when looking at boiling water. In so far as blindsighters are susceptible to visual illusions, they can surely arrive at introspective conceptions of this type. So the set that Sellars' intersubjective predicate 'red' denotes (i.e. the set of perceptual states which have the property expressed by such theory-laden predicate) contains items that lack phenomenality altogether. 'Looks red' or 'phenomenally red' cannot inherit their denotation from a higher-order concept conceived intersubjectively (as referring to a publicly accessible theoretical posit).

As for causal co-variance theories, Leeds finds that they are unsuitable for phenomenal concepts (see, for example, Fodor's (1987), where one such theory is advanced). He observes that causal covariance allows for misapplications of concepts, that is, they allow for the possibility that my tokening of 'phenomenally red' is occasionally brought about by, say, a green quale and therefore do not capture one

peculiarity of phenomenal conceptions – that they are used with certainty and that phenomenal beliefs are infallible.²⁵

For, if P is whatever property has caused my past utterances of ‘Something is P’, then I ought to at least leave open the possibility that there is no one such property, or that the causal mechanism has somehow gone haywire (Leeds, 1993, p.320).

“P is whatever property has caused my past utterances of ‘Something is P’” is not to imply that all possible external causes of my application of P will be included in its extension. There is a method of differentiating between causes which are not individuating of P’s meaning and those which are in non-semantic and non-intentional terms.

For instance, in the work mentioned above, Fodor offers an explanation, in purely causal terms, of why the extension of ‘horse’ remains stable despite the fact that on some occasions ‘horse’ tokenings are caused by cows. What secures that cows do not become a part of the extension of ‘horse’ is that the causal connection between cows and ‘horse’s is asymmetrically dependent on the causal connection between horses and ‘horse’s. The asymmetry is captured in the following: 1.) A’s cause ‘A’s. 2.) ‘A’ tokens are not caused by B’s in nearby worlds in which A’s don’t cause ‘A’s. 3.) A’s cause ‘A’s in nearby worlds in which B’s don’t cause ‘A’s. Nevertheless, the possibility, however faint, that connections between extralinguistic causes and concepts can somehow be rewired remains. Accepting such consequence amounts to a denial that phenomenal concepts exist.

Having rejected causal co-variance accounts as well as the proposal that phenomenal concepts are purely recognitional, Leeds’ solution is a partial identification of phenomenal beliefs with sense impressions. It is partial in that sense impressions and

²⁵ Compare Brandom’s diagnosis of the ostensible infallibility (or incorrigibility) of phenomenal beliefs based on Sellars’ understanding of ‘looks-F’ as a withdrawal of one’s endorsement of ‘is-F’: “There is no sensible contrast between looks-to-look F’ and ‘looks-F’, of the sort there is between ‘looks-F’ and ‘(is-)F’ because the first ‘looks’ has already withheld endorsement from the only content in the vicinity to which one might be committed (to something’s being F). There is no further withholding work for the second ‘looks’ to do. There is nothing left to take back. Since asserting ‘X looks F’ is not undertaking a propositionally contentful commitment – but only expressing an overrideable disposition to do so – there is no issue as to whether or not that commitment (which one?) is correct.” (Sellars, 1997)

phenomenal beliefs are still conceived as discrete representations. The latter, however, are assimilated to the former in the sense that both are parts of the output of the visual system. Phenomenal beliefs are formed in the same mechanical and involuntary manner as sensory images, and are subsequently fed into the belief box. As pre-linguistic infants have sense impressions, they also have phenomenal concepts at their disposal virtually from birth.

What is it about these innate representations that makes them belong to the intellectual order? It is their potential, by virtue of their syntactic structure or 'shape', to causally interact in inferential transitions with concepts added to one's cognitive architecture on acquiring a language. Implicit in the syntactic shape of a native concept are all possible interactions with what is added through learning later in life. Sense impressions do not of course have the kind of structure that would enable their contents to be transferred to the epistemic level directly. Their contents must therefore be captured by representations which are apt to play a role in thought processes, should one's cognition contain a sufficiently large battery of concepts for thought processes to take place.

So I take the gist of the proposal to be this: We may agree with Sellars that public language fails to deliver concepts which apply exclusively to phenomenal properties precisely because intersubjectivity is built into introspective conceptions regarding perceptual states and properties (recall that 'looks red' is parasitic on 'is red', and that sense impressions are publicly available theoretical postulates). But all concepts that are supplied in the learning of a public language must meet certain structural or syntactic conditions in order to have the causal efficacy, within our cognition, distinctive of epistemic items. Leeds would see no reason why innate concepts cannot satisfy such conditions; and the above assimilation of phenomenal beliefs to sense impressions is to guarantee that the denotations of phenomenal concepts do not include non-qualitative perceptions.

Leeds now proceeds to demonstrate that his innateness hypothesis is reconcilable with justificational holism. The epistemic components of the outputs of the visual system do not immediately function as beliefs in their pre-linguistic possessors. The rather ingenious suggestion is that phenomenal concepts lie dormant, as it were, in pre-linguistic

subjects. Consider, for example, the sentence 'Ran' (a is R now) where R is a phenomenal predicate:

Indeed, although the inner state S which is the assertion of 'Ran' is, I am suggesting, available from the beginning, our justification for taking S as the assertion of a particular sentence with a particular form is based on inferential roles which S may come to play only much later in the child's development: for example, it is only when the child has gotten in the habit of responding to occurrences of S by storing in memory beliefs which we can interpret as having the form (s)(t)Rst (i.e. the child remembers having had a 'red' experience, but has forgotten when or in which part of the visual field) that we can look back and interpret S as an assertion of 'Ran'. This is the remark, promised earlier, which I hope makes the innateness claim a little more acceptable: a child is not born having beliefs; he does however enter into states which will later play the roles of beliefs. (Leeds, 1993, p.391)

Leeds himself sees what he in fact offers as a use theory of meaning for phenomenal concepts. But whether this does not violate Sellars' version of inferential role semantics (or justificational holism), however, remains to be seen. Some quite disparate positions have been lumped under the label 'use theory of meaning', so it is paramount, in assessing its compatibility, to have full understanding of the differences between inferential role semantics advanced in the context of a mentalistic approach to intentionality, and inferential role semantics advanced in the context of a non-mentalistic approach to intentionality. I already touched on the contrast between mentalistic and non-mentalistic accounts near the beginning of 2.2.4, but the present matter calls for a more detailed characterisation. Plainly, Leeds' innateness hypothesis belongs to the former category, whereas Sellars' contention that inner episodes such as thoughts and beliefs are internalised public-language sentences fall into the latter.

The relevant differences are brought out by considerations regarding dependence-relations between thoughts and public language. One can be said to have priority over the other in the following respects (I borrow these from Davies (1998)): ontologically,

epistemologically and analytically. In what priority orders do thoughts and language come, according to those three criteria, within a mentalistic framework? A ‘use’ theory of meaning for Mentalese concepts in fact comes down to a functional/computational role theory of the content of mental representations, hence it is easy to anticipate what the respective answers are. The Mentalese hypothesis is a commitment to a self-contained content-generating network of representations which is ontologically, epistemologically and analytically independent of public language, even though a substantial number of representations may be added to the network upon learning a language (Fodor (1975), for instance, maintains that all concepts in *Lingua Mentis* are innate and that its vocabulary is as large as the vocabulary of English, but most functionalists working outside MIT would not feel compelled to make such a strong innateness claim). On the other hand language is, from a functionalist point of view, entirely dependent on thoughts in all three respects – there can be no language without thought, public-language concepts owe their semantic properties to the contents of thoughts, and finally content is a more basic notion than meaning in order of analysis.

What orders of priority does the Sellarsian standpoint entail? We saw in the previous sub-section that his view of the relation between thoughts and language is, in essence, that

...the ability to have thoughts is acquired in the process of acquiring overt speech and that only after overt speech is well established, can “inner speech” occur without its overt culmination” (Sellars, 1997, p.105).

As for ontological order of priority, note that on a non-mentalistic approach the relation cannot be the reverse of what it is if the truth of a mentalistic approach is assumed. There must be an inner causal root of every verbal performance. Overt speech cannot occur without being rooted in an inner episode. Hence the relation is, rather, that of mutual dependence. What a non-mentalistic approach does reverse in comparison with the Mentalese hypothesis, however, are epistemological and analytical orders of priority – the semantic properties of verbal performances are carried over to inner episodes (i.e.

thoughts), which implies that meaning is a more basic notion than content in order of analysis.

Leeds' contention is that his innateness claim would be considered somewhat less objectionable from Sellars' perspective due to the implication that an output of the visual system does not function as a concept and does not acquire its semantics until it comes to play a certain role in inferences. The inferential potential of the output is implicit in its syntactic shape or 'spelling' even prior to its doing so. But bearing in mind the contrast between this proposal and Sellars' version of inferential role semantics regarding orders of priority, it follows from the latter that, firstly, content does not supervene on functional organisation (or functional organisation plus an external causal condition), and secondly, that a representation's inferential role cannot be implicit in its shape or structure when considered in isolation.

Let me begin with the first point. To give linguistic meaning explanatory priority over content in the way that Sellars does is to say that inferential patterns are established in verbal communication, by the practice of giving and asking for reasons in particular, not by a self-contained and isolated causal network of representations which can be fully in place regardless of whether one is a speaker of a language. Now, granting that inferential patterns are formed by communicative practice, could it be replied that the inner representations which public language delivers may have a structure by virtue of which they will causally interact with native phenomenal concepts?

Seeing why the answer to this in the negative will bring us to the second point, namely, that if a non-mentalistic account is true, the structure or shape of inner representations is not such that if we focus purely on those intrinsic properties, we will be able to anticipate what thought processes they will appear in. I think we only need to draw attention to certain differences between public language and Mentalese to show that the response is not available. Although Mentalese is thought of as a language, it is a highly idealised language in that each representation has a unique spelling or shape. On the other hand, one English word can express two or more concepts, e.g., 'conductor' meaning a substance that allows electricity to pass through it as well as a person who directs orchestral performances. In the case of Mentalese, the distinction between words and concepts collapses – these two meanings would be assigned to representations with

different spellings, say, ‘conductor₁’ and ‘conductor₂’ (but we need not suppose that there are many similarities between their spellings). Because in public language the difference between the two concepts is not reflected in the phonetic features of the words that express them, the only way to reveal the different concepts is by looking at entire inferential patterns formed by sequences of verbal performances containing the word. Likewise, all inner correlates of the word ‘conductor’ will be alike in their intrinsic properties no matter whether on one occasion it is used with its first meaning and its second meaning on another. It is for this reason that a non-mentalistic account of this sort takes inference and not a single proposition as the basic vehicle of content.

The upshot then is that Sellars’ non-mentalistic approach rules out the possibility that a certain class of concepts, however small, can be innate. Since two or more inner representations with different meanings can be indistinguishable in their syntactic features, no determinate meaning can be assigned to ‘concepts’ that are available from birth. Should beliefs understood as internalised public-language sentences ever engage such innate items in inferential transitions, there is no guarantee that they will be engaged as *phenomenal* concepts at all. So even though Leeds’ idea that an output of the visual system does not gain an epistemic status until it comes to play a role in inferences is a much less controversial version of the innateness hypothesis than Fodor’s, he has not shown that it is compatible with Sellars’ inferential role semantics.

The final point to be made in this section concerns the mention of normativity in the claim I have been making a case for, namely, that

(iii) to have a higher-order thought about one’s perceptual state is to token an internalised public-language sentence whose meaning is holistic and normative.

I have already noted that if public language is viewed as the medium of thinking, inferential patterns are needed to assign different concepts (and hence different meanings) to certain phonetically identical expressions on the one hand, and to make certain pairs of phonetically different expressions synonymous on the other. That a non-mentalistic treatment of the sort I have been discussing is bound to involve normativity – whereas a mentalistic one (i.e. a functional/computational role theory) isn’t – is an immediate consequence of the fact that inferential relations between verbal performances

are not causal (i.e. non-normative) in the way that inferential relations between representations realised in the brain are. It would not make much sense to say that inferences are computations over utterances, and so understanding one's assertion, on a non-mentalistic approach, amounts to knowing what set of premises *entitle* him to it, and what consequences it *commits* him to. Functional/computational role semantics dispenses with such normative notions entirely.

So much for (iii). We need not go further into the issue of normativity as the ideas that guide my argumentation towards (iv), which was the claim that

(iv) since higher-order perceptual thoughts are internalised public-language sentences, qualia are intrinsic properties of experiences,

have already been set out. My aim in the next and final section of this Chapter is to show that with Sellars' view of introspection in place, we should feel drawn towards (iv) and even further until phenomenal individuals such as sense-data will make another appearance. Given that in Chapter 1. I undertook to gradually shift qualia from phenomenal individuals to wide contents, where the first step consisted in dissolving phenomenal individuals with the aid of a Lycan-style adverbial analysis, their second appearance will be rather unwelcome. I will of course dedicate Chapter III. to exploring our prospects for an alternative representational account, one which will not force on us the epistemologically dubious purely recognitional concepts or innate Mentalese concepts.

2.3 Sensa and Phenomenal Individuals

The discussion in the last four sub-sections was dominated by a search for a conceptual difference at the higher-order level between individuals with inner life and those who, at least under certain circumstances, are perceptually sensitive but don't experience anything. It was the failure to find one (conceptual or other) at the first-order level that brought higher-order concepts into focus. The implications reached beyond

first-order representationalism, for whatever first-order theory one is prepared to back, it will at best account for what normally sighted people share with perceivers who don't experience – worldly subjectivity rather than experiential subjectivity (I of course spent much of Chapter I. defending FOR against some of its main competitors).

The discussion of the nature of introspective awareness was to bear directly on the tenability of higher-order representationalist accounts of qualia. The view adopted, on which awareness of experiential inner episodes and their properties is the result of considerable conceptual sophistication, left no room for the two kinds of concepts that could sustain HOR, namely, purely recognitional concepts and innate Mentalese concepts. Neither brute acquaintance nor tying concepts to the outputs of sensory processing almost to the point of assimilation have a part to play in that process of conceptual sophistication.

The implications of the challenge I present for HOR are best brought out against the background of Sellars' remarks about what an advanced science of perception will have to say about phenomenal properties, what phenomenal properties turn out to be within the so-called Scientific Image. The Manifest Image, on the other hand, was to explain (as we saw in 2.2.4) how we ordinarily come to understand each other and, more importantly, ourselves as creatures with inner life. The conceptual development within the Manifest Image eventuated in non-inferential/observational awareness of feelings and sensations (recall that at the initial stage we were all 'behaviourists' – there was no awareness of anything occurrent).

Now, what is the difference between the status of phenomenal properties within the Manifest Image and their status within the Scientific Image? Let me begin with the first part of the question. It is to ask, in effect, what our folk-psychological theory of our own mind (whose development Sellars traces) says about phenomenal properties. The most important point that follows here is that in the Manifest Image only worldly objects are coloured, nothing inside is. Experiences, which we come to be aware of at some point in infancy, are conceptualised as states (whether or not explicitly). These are not coloured, although they, again,

...stand to one another in a system of ways of resembling and differing which is structurally similar to the ways in which the colours and shapes of visible objects resemble and differ (Sellars, 1997, p. 112).

The colours of worldly objects that Sellars speaks of, however, are not the colours of naïve realism (intended to be a branch of phenomenalism), for the is/seems distinction continues to apply in their case via the introduction of standard conditions, with the notion of standard conditions, to repeat what was stressed in 2.2.4, being woven into the Manifest Image conceptions in question. This is certainly true, as ordinary talk of colours frequently makes use of the distinction. To follow naïve realists in collapsing the distinction is to stray unreasonably from what is characteristic of the Manifest Image conceptions.

It is only in the Scientific Image that we can coherently claim that nothing in the outside world is coloured, let alone coloured in the naïve realist's sense. The objects of the common-sense world are made up of colourless parts. The middle-sized objects of the common-sense world that we perceive as being homogeneously coloured are in fact collections of microphysical entities. Nothing in the world has the sensible properties that common sense attributes to objects as they are experienced, which

...suggests that in the scientific picture of the world the counterparts of the colours of the physical object framework will turn out to be aspects, in some sense, of the percipient organism (Sellars, 1963, p. 99).

It is Sellars' contention (not entirely unquestionable) that by running counter to our pre-theoretic framework of middle-sized objects and persons in the world, physics abandons the framework of macro-physical objects altogether to the effect that only micro-physical entities (molecules, electrons, etc.) really exist. Only macro-physical objects could have the properties they are perceived as having, but according to Sellars, the Scientific Image rules out their reality.

It is Sellars' next move that I am keen to connect with the key points in my challenge for representational theories of qualia. When explaining what becomes of perceptual states, or sense-impressions, in the scientific picture, whether it can preserve them, he says that sense impressions are states of persons which in the Manifest Image are single logical subjects. Sense-impressions are folk-psychological posits on which our introspective understanding of the qualitative part of our mental lives is based. However, as was the case with macro-physical objects, the scientific picture recognises no such thing as a single subject of perceptual states. Like macro-physical objects, persons are swarms of micro-physical parts which cannot serve as subjects of perceptual states. Science once again contradicts the truths in the Manifest Image, in this case the common-sense framework of persons. A functionalist would of course object that the self is not a matter of micro-physical constitution and that it is at the functional level of description of an organism that the Manifest Image framework of persons is vindicated. Sellars anticipates this and considers an electronic robot which is sensitive to its computational processes in the language of persons. An organism's functional organisation is a system of interacting sub-personal components such as the mind-reading module, the information-processing unit, etc. Sellars finds the following problem:

...the fact that they are a plurality precludes them from serving either jointly or separately as the subjects of the verb 'to sense red-triangle-wise', We must therefore either introduce another logical subject (an immaterial substance) to do this work, or turn each sensing into a logical subject in its own right, i.e. introduce a new category of entity ('phantasms' or 'sensa' we might call them) with predicates the logical space of which is modelled on that of visual impressions, as the latter was modelled on that of coloured and shaped physical objects (1963, p. 101)

The first step removed macro-physical objects from the Scientific Image, the second one does the same with selves. Perceptual states will also be discarded as a result of these moves, for there are no unified persons which perceptions are states of. Since the ultimate account of the world does not accommodate persons, states are succeeded by individuals and phenomenal properties are properties of individuals rather than states.

To begin, it must be pointed out that the representationalist has the most solid grounds of all relationalists for opposing the idea that there is a tension between the scientific claim that only colourless micro-physical entities exist (and hence that macro-physical objects don't), and the common-sense attribution of colours to macro-physical objects. The objects of the common-sense world and the properties they are perceived as having (qualia) should be treated as intentional, he suggests, which is consistent with saying that only colourless molecules and atoms exist. He would perhaps think it encouraging that his position reconciles the two Images by giving a naturalistic account of non-conceptual intentionality involved in experiencing in terms of some nomic relation between micro-physical objects and sensory states of organisms, for example. Does this mean that every representation is a misrepresentation given that nothing in the physically described world matches the non-particulate character of macro-objects as presented in experience? The representationalist only needs to add that under normal conditions of perceiving colour is at least partly determined by certain features of micro-physical objects (whether they absorb or reflect light in particular) and that these features produce representations of homogeneously coloured (i.e. non-particulate) objects, and this is a fact about the design of the visual system, not a troubling failure to see the world as it is.

Having said this, representationalism is powerless to reconcile the two Images when it comes to persons. If there isn't a single subject of perceptual states, constructed out of interactions among the components of a computational systems or otherwise, then experiences are not states at all, and will be replaced by phenomenal individuals. This follows regardless of whether the claim that qualia are intentional is consistent with the Scientific Image.

Here are the points where Sellars' reasoning about the subject of experience connects with my argument in 2.2.1 (*Subject-Dependence of Qualia*) which developed a great deal of the challenge that I am now about to conclude. My thought, inspired by Dennett's (1978) project of building the self out of sub-personal material and repeated a number of times in what followed, was that phenomenal properties and single subjects are inextricable, that the self is somewhat immersed in them, and that separating them would either change the topic beyond recognition or lead to phenomenal individuals. More specifically, phenomenal individuals will be unavoidable if the self is separated from

qualia because this turns the self into an observer in relation to qualia, which then turns qualia into properties of phenomenal individuals. The self becomes an inner sense capable of directing and opening itself to phenomenal objects. If states rather than objects are what carries phenomenal properties, the self must be related to them as their subject, not as observers. In view of this, Sellars is right in thinking that in the absence of a single subject we lose perceptual *states* too.

This was aimed at strictly first-order theories and was to provide further support for Carruthers' argument that first-order representationalism is an account of worldly subjectivity shared by unconscious perceivers. To explain experiential subjectivity, where inner life is to be found, it ought to involve self-reflexive states. By 'strictly first-order' I mean those positions on which qualia are not only individuated but also fully determined at the first-order level. Failing to appreciate the importance of the subject, experiential subjectivity would prove elusive. I wasn't quite prepared to go along with Sellars in dismissing the ordinary framework of persons from some scientific point of view at the time. I did not comment much on the undeniably difficult problem of subjectivity and thought instead that concentrating on the nature of self-reflexive states promised more tangible progress. After all, it seemed sensible to postpone going into the issue of the self, since I believed, as I still do, that because higher-order representationalist proposals do build self-reflexive states into phenomenal consciousness, they may be thought of as recognition of the fact that qualia and the self are inseparable. Bringing introspection into focus was likely to be more fruitful because discovering that self-reflection features phenomenal concepts, i.e. concepts not shared by unconscious perceivers, would serve as evidence that the source of such self-reflexive thoughts is the kind of self that is 'immersed' in phenomenal properties.

The outcome should by now be largely clear. Introspective concepts do not distinguish us from perceivers who don't experience. For various reasons, many would not want to halt all progress with the problem of the subject of experience in the scientific picture quite as resolutely as Sellars does, but the alternative route via the examination of introspective concepts appears to reinstate objects at the expense of states with equal force. Let me explain this in more detail. In section 2.2.1 I also outlined a general form of higher-order representationalism. In order to place sufficient emphasis on the role of the

subject and to be regarded as an account of experiential subjectivity, it must meet the minimum requirement that lower-order states be available to be picked out by higher-order states. The result is a view where qualia are partly determined by long-armed functional roles and partly by being related to the subject in terms of availability to introspection.

What position succeeds HOR given that introspective states are internalised public-language sentences containing nothing that applies distinctively to phenomenal properties? It will not help to return to the first-order level for we would repeat the mistake of misidentifying phenomenal consciousness with worldly subjectivity. It seems that the only option now is a position which is neither first- nor higher-order, one which abandons a reference to states altogether, thereby removing the distinction between perception and introspection. That position is the sense-datum theory according to which experiences are sensings of phenomenal objects, i.e. sense-data. Why can it not be characterised as either first- or second-order? Why should it be understood as assimilating perception to introspection? These questions can be answered together. If we take perceptions as well as introspecting to be states (as of course HOR does, for example), then perceptions are first-order states, whereas introspecting is to be in a higher-order state. At the beginning of Chapter 2. I grappled with difficulty of whether the sense-datum theory ought to be understood as a higher-order position. It is its rather anomalous nature that causes the difficulty. It is correct, on the one hand, to suppose that higher-order awareness is awareness of the contents of one's mind no matter whether these are perceptual states or phenomenal objects, in which respect the sense-datum theory is a higher-order theory, and to also suppose, on the other hand, that there can only be higher-order awareness where there are first-order states, in which respect awareness of sense-data is first-order awareness. I think the best way to clarify the situation is by saying that it is neither because on the sense-datum theory perception and introspection are one and the same thing. So what I mean by the sense-datum theory's assimilation of perception to introspection is not that it adopts the standard higher-order theories' layout with first-order perceptions and higher-order introspecting, and has it that introspecting is itself like perceiving rather than an occurrence of a thought.

So the sense-datum theory has none of the drawbacks of first-order or higher-order theories and combines the virtues of both. Most importantly, it accommodates experiential subjectivity without being higher-order, whereby it does not depend on the existence of phenomenal concepts. HOR, whose qualia-determining factors range from external objects to self-reflexive states, implodes into something considerably simpler – experiencing comes down to the sensing of sense-data, to opening the inner sense to phenomenal individuals.

It goes without saying that reintroducing sense-data comes at a great cost not least because such an outcome leaves us at a complete loss as to why sense-data occur in some perceivers but not in others. Even more troubling is the fact that it closes the circular argument that started with Lycan’s attempt to purge the mind of phenomenal objects by treating after-images, supposedly the prime examples of sense-data, as intentional inexistents. His diagnosis of the perceptual situation involving a green after-image is summarised in the following:

There is a mode of sensing M such that Leopold is sensing in way M and Leopold would be sensing in way M if a green patch were present to him and conditions were normal,

and Lycan adds that

the antecedent of this counterfactual is “there is a green patch present to Leopold”. Thus, any of the standard accounts mentioned above directs us to look at an alternative world in which (it is really true that) there is a green patch present to Leopold. And this is the source of our quantifier: the value of its variable *is* a green – physically green – patch (1987, p.87-88, italics original).

I presented this as a basic sketch around which representationalism can be built and at the end of Chapter I. I arrived at the claim that (i) qualia are not intrinsic properties of perceptual states, and if so, they must be wide contents.

Having reintroduced phenomenal objects in the above way, we encounter the general problem – to be dealt with in the final chapter – that my circular argument raises. Sellars is one of many authors who favour a non-mentalistic approach to the intentionality of thoughts. Thoughts inherit their intentionality from the semantic properties of public language (which implies, as I briefly explained at the end of the previous section, that intentionality is normative). The idea that concepts do not appear out of nowhere, but their grasp presupposes adopting the communicative practices of the linguistic community of which one is a member was at the heart of my criticism of phenomenal concepts. That intentionality is normative and therefore irreducible is not something we should find problematic. It is not a concession to dualism, as normative phenomena are not targets for reduction. The issue my circular argument raises is that the normativity of introspective thoughts does become problematic if they turn out to be indispensable in an account of a phenomenon that is not a matter of collective conventions and hence *is* a target for reduction. That phenomenon is of course experiential subjectivity or phenomenal consciousness. Wide content representationalism is the strongest reductive strategy we currently have, or so I argued. As an account of experiential subjectivity, it cannot dispense with something which, unlike the visual system, is not part of our phylogenetic equipment. Finally, accepting that it collapses into a position which characterises phenomenal properties as properties of objects floating in the mind is a substantial concession to dualism.

We have now reached the last part of my work where I want propose a representationalist view which I believe stands a chance of addressing the problem by tying qualia even closer to the subject than the standard form of higher-order representationalism I have been discussing. The proposal is that qualia are properties of higher-order states.

3. An Alternative Proposal

3.1 *Qualia as Properties of Global Higher-Order States*

In presenting the position that deals with the difficulties we encountered in Chapter 2., I shall first bring together the main themes developed in the course characterising representationalism in Chapter 1., as what I am about to put forward adopts the understanding of content externalism we arrived at. It was mainly the clarification of the notion of a functional profile and that of a long-armed functional role (especially 1.2.1 *Functionalism and Intentional Contents*) that shaped that understanding. They are of course interrelated – functional profile is just the totality of the functional roles of the states a system can enter in the context of what their outputs are supplied for (e.g. cognition, motor control, etc.). Functional profile also explained the attribution of less determinate contents to visual systems with poorer sensitivity. If an organism is unable to make out certain shades of colours, its functional profile subsumes two or more external inputs under the same output. We don't need make a firm decision as to whether the causal links between the environment and states of visual systems should be understood in terms of causal co-variance or teleologically, although some of my defence of content externalism in Chapter 1. did rely on anchoring contents in the environment teleologically.

On the view I have set out to outline, these externally determined contents are non-qualitative, but I am keen to preserve the above points about content as well as everything else that was said about them in Chapter 1. I want to add, however, that the intentional objects of first-order states have their properties homogeneously. This party reiterates the insight offered by representationalism in response to Sellars' contention that there is a friction between the Manifest Image and the Scientific Image in that nothing in the world has the homogeneous colours that common sense attributes to external objects since only colourless microphysical entities exist. I replied on the representationalist's behalf that this is an oversight of the possibility that the colour terms of common-sense talk used to refer to the homogeneous colours of macro-physical objects express the properties of intentional objects. My current suggestion is in agreement with this move to the extent

that it is right to treat perceptions as contentful states and that the contents in question are properties which represented objects possess homogeneously. The general picture here is that the impact of heterogeneous reality on the visual system produces representations of macro-objects with homogeneous features, which features are categorised by long-armed functional roles. All I disagree with is that the features are phenomenal.

Keeping a firm grip on the ontological status of intentional objects and properties (particularly in the context of non-conceptual representations) was also a matter which received careful attention near the beginning of Chapter 1. Inadvertent confusion of the causal powers of intentional objects and properties with those of physical things in the proper sense of the term can result in the mixing of relations which only hold at, say, the personal level with the underlying sub-personal relations, not to mention the blurring of the distinction between functional and structural levels of descriptions or Sellars' Manifest and Scientific frameworks, thus turning a potentially fruitful set of ideas into something that is impossible to evaluate.

Assuming the truth of content externalism, the intentional object of a non-conceptual mental representation is, trivially, topic-neutral. Its topic-neutrality, however, is to be distinguished from the topic-neutrality of the content-bearing state. Content externalism is a broadly functionalist approach and it is a familiar point that functionalism entails no ontological commitments despite the fact that functional states must ultimately turn out to be realised either materially or immaterially. Further, the states themselves have a functional reality, they exist only relative to a system, which is not to introduce a new kind of substance, some functional 'stuff' that emerges from matter organised in a certain way – abstracting away from the states' specific type of (presumably material) realisation simply has the purpose of facilitating formulation of law-like generalisations and prediction of the behaviour of the whole system. But the causal relations which hold the system together are nonetheless relations among physical things in the proper sense of the term. Incidentally, Sellars' insistence that only micro-physical rather than macro-physical objects exist is driven by a concern to steer clear of issues to do with property-emergence.

On the other hand, the intentional objects of the non-conceptual mental representations we are considering are topic-neutral and hence, again, there is no entailment of an ontological commitment, but questions about what they will ultimately

turn out to be are clearly misguided. They are neither material nor immaterial. Further clarification of their status, while keeping the functional and external causal condition in view, will emphasise their subjective reality. Intentional objects exist relative to the subject (e.g. Dennett's 'exempt user'), something that is external to the visual sub-system and exploits the outputs of that sub-system in ways which are sensitive to their intentional objects. 'Sensitivity to content' means that the intentional content is preserved through the course of the processes in which the user involves the corresponding representation and acts with respect to that content even though, for example, its external cause as it occurs does not match the content (and hence fails to satisfy it), or its external cause occurred in distant past (if the representation is retrieved from memory, albeit in a less vivid and degraded form). The most basic requirement for something to serve as an exempt user is that whatever it is that is sensitive to an output's intentional object (which stands, as it were, in an external cause's stead) is the same as that which coordinates motor responses with respect to the object.

We therefore must bear in mind that this does not introduce yet another substance – some subjective 'stuff' – from which intentional objects are made, and that the representation has these causal powers by virtue of its physical realisation. At any rate, talk of subjective reality of intentional objects can only be justified if the outputs of sensory sub-systems are properly related to the exempt user.

While my claim is that first-order intentional objects as I have described them lack phenomenal properties, these remarks constitute the first important step towards phenomenal consciousness. The idea that underlies my remarks about subjective reality is that only physical things have objective reality, but this is by no means intended to make our talk of subjective reality appear derisory. All I aim to show is that if matter (in this case the brain) is organised in a certain way, more specifically, if it is arranged into a functional system, it can give rise to phenomena (i.e. intentional objects) which have the same sort of significance for that system as things that have objective reality without postulating new kinds of substance existing alongside material things. For these very reasons, the form of functional arrangement of matter that I believe gives rise to experiences will do justice to phenomenal properties' subjective reality without the contrast between subjective and objective reality becoming an ontological one. Of course

many functionalists, and representationalists in particular, have attempted to do just that, and yet as many so-called qualia realists remain unconvinced. Qualia realists' objections to every functionalist theory that has been advanced are based on what has come to be known as 'qualia intuitions' – they draw their force from reflections on the supposedly occurrent nature of phenomenal colours and the fact that they are presented in experience as occupying more or less clearly defined regions of space or, alternatively, as saturating portions of one's visual field. Such evidently non-relational and non-dispositional occurrences could not possibly be generated by functional systems, or so we learn. I think it is correct to say that the idea of homogeneity is what unifies these intuitions, that homogeneity lies at their root. I will argue that the account I am proposing allows me to probe further into the intuitions than perhaps most relationalist positions do until they are shown to be futile as a challenge.

Note that nothing in the qualia intuitions will prompt a qualia realist to deny that first-order intentional objects can have their properties (categorised by long-armed functional roles) homogeneously precisely because I do not claim that the properties are qualitative. That a certain functional organisation generates homogeneity would be seen as unproblematic from a qualia realist's point of view provided that the homogeneity in question is not phenomenal. It is only homogeneity which makes itself felt in some way or other that is supposed to be irreducible to functional roles. Separating homogeneity from phenomenality, although we cannot of course have the latter without the former, clears the way for a position on which qualia are the contents of higher-order states representing the non-phenomenal (but sensory nonetheless) homogeneity of first-order intentional objects. The original objection wielded against functionalism was that functional roles could not produce the kind of homogeneity we experience, but my suggestion is that experienced homogeneity develops from homogeneity that we already have in place in the form of first-order contents, one which qualia realists should not find problematic. This has an indisputable potential to silence qualia realists, or at least the burden of explaining why non-phenomenal homogeneity is a problem for functionalism is certainly on them. It is difficult to see, in any case, what explanation qualia intuitions alone would motivate. The main strength of my suggestion consists in placing an

intermediate step between functional roles and experienced homogeneity, namely, non-phenomenal homogeneous intentional objects.

More clarification will be provided, however, as I further develop my account. My immediate interest now is in introducing the idea of a global non-epistemic higher-order state. A global non-epistemic meta-representation is a representation of the contents of the outputs supplied by all five sensory modalities as well as proprioception at any given time. It is global in that it integrates all perceptible features of my surroundings as well as all sensible states of my bodily parts into a single content. Due to its totality, i.e. its global scope, it is subject to continuous changes as, for instance, some of my current bodily discomforts become less severe, I scratch an itch on my left palm, a new scene comes into my view when I look in a different direction and so on. My view then is that qualia such as phenomenal colours, sounds, smells, twinges, throbs, cramps, awareness of the position of my toes, the feeling of my lungs expanding as I breathe in, etc. are portions of the content of a single meta-representational state a perceiver happens to be in at a given time.

In defending the identification of qualia with global higher-order contents I shall seek to explain what it is that sets a global higher-order content apart from first-order contents and makes it qualitative. I shall also finalise my reply to those who are sceptical about relationalist views in general on the grounds that qualia are experienced as being occurrent rather than relational (and that it is self-evident that for a colour to be experienced is for it to be occurrent) by explaining how something can be occurrent from the subject's perspective despite its relational nature. Let me begin by highlighting the key differences between first- and higher-order (global) contents. First and foremost, the contentful outputs of sensory modalities are utilised as states of *discrete* sub-systems. I have stressed that talk of content in this case is justified only if these outputs are related to an exempt user. We can think of the exempt user as some sort of control unit which coordinates the movements of the limbs in response to such contents according to the current needs and motivations of the whole system. In section 2.1 (*Worldly vs. Experiential Subjectivity*) I rehearsed a number of examples mentioned by Carruthers of how non-qualitative perceptual contents can be used to guide swift and instinctive motor reactions, not to mention blindsight.

The point to note is that whatever serves as the user of first-order contents (and whatever the exact manner of their utilisation is), it is external to the sub-system that supplies them. A first-order perception does its representing for something other than the sub-system itself. I deliberately avoid speaking of persons as being the users of these first-order perceptual states to emphasise that the problem of identifying a single subject of perceptual states does not arise if they are non-qualitative, for it arises distinctively for experiences. We can grant that the motor-control unit is a sub-personal part, although for the states to be contentful it must coordinate motor responses in accordance with the needs and motivations of the *whole system*.

On the other hand, a global meta-representation is a single self-directed state aimed at the first-order contents of all sensory sub-systems – and the content of my global higher-order state is, unlike the contents of first-order states, qualitative because the situation now is not such that the output of a discrete sub-system does its representing of something red *for me*, it is, rather, that *I* represent something red. I believe that only the latter ('I represent something red') does full justice to the subjectivity of qualia and that the correct statement of every experiential situation should have this form. For I am claiming, in effect, that the global higher-order state is what the experiencing self comes down to. I am experiencing something red when looking at a ripe tomato because the experienced redness is integrated into the higher-order content of a state which also represents the position and state of my body and its parts and everything else that there is to be sensed and perceived at this particular moment. In short, it becomes part of the content of a state which represents everything that constitutes me as I experience myself.

We may struggle to understand why there should be a certain feel to the content of a representation produced by a sub-system whose receptors are sensitive to sound waves, and a certain (other) feel to the content of a representation produced by a sub-system whose receptors are sensitive to light, especially when these sub-systems produce their representations in isolation from one another. But we will not need to strain our powers of understanding quite so much if the suggestion is that both experienced colour and experienced sound are parts of the content of *the same personal-level* representation aimed at first order intentional objects. Having said this, it remains the case that the

differences between visual and auditory qualia are determined by the character of the first-order intentional objects picked out by the meta-representation.

An all-encompassing meta-representation is presumably a brain state and I have no reason to deny that at one level of description it is sub-personal, although it is equally legitimate to describe it as a personal-level representation. Nonetheless, what matters for my purposes is that owing to the global scope of the meta-representation, there is no such level of description at which its content is sub-personal. This overcomes the difficulty discussed in the final section of the preceding chapter that draws Sellars towards phenomenal individuals. His reasoning was that sense impressions are states of unified subjects which have no place within the Scientific image, since living organisms, including humans, are bundles of countless micro-physical particles. Even attempts to construct the self out of interactions among the sub-personal components of a computing machine inevitably fail, as we still have a plurality instead of unity. Because sense impressions depend on the existence of a single subject, we are forced to replace states with objects as the bearers of phenomenal properties. Sellars' observation that sense impressions require a single subject whose states they are is not something to be concerned about if we do not hold that (first-order) perceptual states are the bearers of phenomenal properties. My account only requires that the states have a wide content and hence satisfy the conditions mentioned above (long-armed functional roles and the presence of a motor-control unit serving the system as a whole). Clearly, if first-order perceptions are not experiences, we are under no obligation to identify the experiencing self.

So I avoid introducing phenomenal objects by claiming that qualia are properties of a higher-order state rather than properties of first-order perceptions. In view of this, the issue of finding a unified subject of what I take to be a qualia-bearing state poses no threat to my position, since the global higher-order state is what the experiencing self comes down to. It is by virtue of being in a such a self-directed global state that the description 'I represent something red' is applicable to my present experiential situation as I, for example, stare at a ripe tomato. The unity, as opposed to multiplicity, is guaranteed by the fact that there is only one such self-directed global state at a time with a single all-embracing content. There is no doubt that the unity understood in terms of a

materially realised meta-representational state is consistent with the Scientific Image plurality regarding sentient organisms.

In 2.2.1 (*Subject-Dependence of Qualia*) I acknowledged that a position on which qualia are first-order contents – provided it relates them to the subject in the required way – is not precluded from moving from ‘There is a representation of something red in me’ to ‘I represent something red’. But this alternative would not fare as well as the view I am putting forward when faced with Sellars’ denial of the existence of a single subject. A direct identification of the subject to which we could relate phenomenal states proved to be beyond reach, as it would endlessly run into to difficulty of having to refer to sub-personal material, thereby inviting obvious counterexamples (which would only confirm Sellars’ doubts). I chose instead an indirect route via phenomenal concepts. Showing that phenomenal concepts were possible would be an encouraging indication that there was an experiencing “I” to pursue after all. At least, quite interestingly, we would be gifted the option of suggesting that there is a global conceptual meta-representation featuring a large but not infinite number of, say, innate mentalese concepts. The state would be far too complex for each concept involved to be applied consciously (i.e. to focus attention on the entire content), but examples such as absent-minded driving are a reminder that non-conscious applications of concepts may often vastly outnumber conscious ones. As for purely recognitional concepts, on the other hand, first-order states must be phenomenal prior to the kind of recognition by means of which one forms a purely recognitional concept, and therefore phenomenal states would still have to be tied to something other than a global meta-representation containing purely recognitional concepts, that is, a subject which the global state itself would be a state of rather than a subject that is reducible to the global state.

Unfortunately, as it turned out, phenomenal concepts are too epistemologically dubious to be any kind of encouragement. I should also add that it would not be open to us to say that all first-order qualitative contents at a given time are simultaneously targeted by a global *non-epistemic* meta-representation. Being a (non-conceptual) representation, its global content would necessarily have its own qualia, thus resulting in an odd duplication where the phenomenal properties of lower-order states are hidden

beneath those of the global one. A lack of its own qualia would mean that we should all be diagnosed, amusingly enough, as ‘sightblinders’.

I hope the basic commitments of my view are now largely clear, and that it is easy to anticipate at least the immediate objections and counterexamples it may prompt. I have yet to spell out how I understand the relation between the global non-epistemic meta-representation and introspection as well as attention, and, crucially, specify where exactly it breaks the circular argument that motivates this solution. But first I would like to return to explaining why qualia are occurrent from the subject’s perspective despite their relational/functional nature. I left off after indicating that we can begin to undermine qualia intuitions by saying that the impact of heterogeneous reality on our senses produces states with homogeneous intentional contents. The thought was that we could rightly expect much hesitation on the qualia realist’s part as to whether he should deny that homogeneous intentional objects can have subjective reality in a functionally described system if we add that they are not phenomenal. For if the point is granted, qualia intuitions will be under considerable strain once we go on to claim that qualia are the contents of a state which represents objects (i.e. first-order intentional objects) that already have some of the features that, according to qualia realists, make phenomenality impossible to capture in relationalist terms.

This argument will only do the work I need it to do when it is conjoined with the second part of my proposal where the idea of a global higher-order state is introduced, otherwise I would not be able to explain why the represented properties of first-order non-phenomenal homogeneous intentional objects aren’t just that – homogeneous and non-phenomenal. Suppose first, for illustration, that I’m viewing an abstract painting depicting a cat in the foreground (suppose there is just the most abstract feline-like outline) and a certain scene in the background. The only occurrent features of the painting are, plainly, the two-dimensional patches of various colours on a large piece of canvas. What it represents, however, is its dispositional feature, one that depends on my understanding of this particular convention of depiction. I mention this to highlight the point that what is dispositional in this example is, *from the subject’s perspective*, in the case of a global meta-representation. This is so because I equate the experiencing self with the global meta-representation. I argued that the identification of the subject with the

global state is justified, on the one hand, by the state's being self-reflexive in that it is directed at the contents of lower-order states, and its global scope on the other, thus ensuring that there is no level of description at which the single global content is sub-personal (although the meta-representation itself is presumably a brain state). Assimilating the subject, the experiencing self, to the global state means that *the subject's perspective amounts to the global state's pointing towards its all-embracing content; and for the subject to be confronted with an analogue (as well as global and self-reflexive) content in this most direct and immediate manner is for that content to make itself felt in a certain way. Needless to say, making itself felt entails being occurrent even though our underlying ontology of qualia is relationalist.*

Whereas in the case of the abstract painting my awareness of its three-dimensional content, i.e. the cat in the foreground and the scene in the background was mediated by awareness of the intrinsic features of the painting, i.e. the two-dimensional patches of paint, the visual portion of the meta-representation's content is presented to me directly and hence in three dimensions – the presented objects have volume and orientation, some are closer than others, etc. This is of course a reference to the familiar idea of transparency, the source of which in the context of on my account is the understanding of the (experiencing) self in terms of a global meta-representation. Since my subjective perspective towards qualia comes down to the meta-representation's pointing towards its content, I have no non-conceptual awareness of its intrinsic features – trivially, its intrinsic features are not part of its content and so are not part of what I have a subjective perspective towards. It now seems right to say that a non-conceptual state to which the statement 'I represent something red' (as I, say, have a ripe tomato in my view) applies without relying on a move from 'There is a representation of something red in me' should be regarded as a paradigmatic case of non-conceptual mental representation.

Qualia realists are unlikely to be wholly convinced and would complain that they still don't find the explanation, as Carruthers likes to put it, 'cognitively satisfying', since something so intensely felt cannot be topic-neutral in the sense that it is neither material nor immaterial. I can only repeat in response that the materially realised global state is already directed at topic-neutral objects (lower-order intentional objects). The metarepresentation 'records' properties of topic-neutral objects, ones which are neither

material nor immaterial. What is more, they encode them as homogeneous – homogeneity, at least in this sense, being a property that has no reality in nature outside sentient organisms – which makes the symbolic structures a unique occurrence perhaps already at the level of biochemistry. Revealing how homogeneity, not only visual one but that of all sensory modalities and proprioception, is ‘recorded’ in the brain would clear the way for an exhaustive explanation of qualitative content. Like all intuitions, qualia intuitions are poorly articulated, yet quite unyielding, hunches that we will have no qualms ditching if such an exhaustive explanation of the mechanisms of ‘encoding’ phenomenal homogeneity can be given.

Similar intuitions regarding non-mental topic-neutral objects prove to be misleading almost without exception and lose all appeal as soon as the ‘realist’ about such objects comes to have a full grasp of, say, the relevant laws of optics. Holography illustrates this well. Holograms are images replicating spatial relations in the original scene along all three dimensions. Take, for example, a sizeable hologram (1m^2) where a Rubik’s Cube appears to be positioned half a metre behind its two-dimensional surface. There is also a smaller cube concealed by the larger one when the viewer aligns himself with them. The image is three-dimensional in that when the viewer moves right or left to change perspective, he can see the side of the large cube and the smaller one behind it. Having no understanding of the methods used to create holographic images, our viewer is likely to be gripped by an intuition as unquestionable on the face of it as those about qualia. He thinks it inconceivable that the cubes he can view from different perspectives are topic-neutral, that there is no cube-shaped substance (material or immaterial) at that precise location.

While a photographic film is only sensitive to variations in intensity, a holographic plate also detects differences in phase between light waves reflected off the cubes being recorded. Differences in phase correspond to relative distances of the objects’ exposed surfaces. Phase is made detectable (in the so-called transmission type of equipment set-up) when the light crosses another beam, one that avoids striking the cubes (so all of its light waves are in the same phase), and hits the plate at an angle. Where the uniform beam and the light waves in different phases travelling from the cubes meet, an interference pattern forms marking the waves’ phases (and hence the relative distances of

the cubes' surfaces). The pattern then leaves a recording on the holographic plate later developed into a hologram.

To view the hologram, the uniform beam must illuminate its back ('back' in relation to the viewer) surface at the same angle that it did during recording. When these light waves travelling in the same phase pass through the image, the optical features of the image obtained by developing the 'recordings' on the plate convert them into so called spherical front waves which leave the front surface of the hologram and spread towards the viewer. Spherical front waves spread in the shape of concentric circles in a ripple-like way. Finally, although the circles begin at the surface of the hologram, their size is such that their centres are behind the hologram, at the very same location where the exposed parts of the cubes that reflected light onto the plate were positioned. In other words, the centres are where the sources of the ripples would be if no hologram was involved and they had spread from point sources. Vision then reconstructs the point source from the wave even though it does not originate in a point source. Hence the point source is virtual or, to place it in the context of my account of qualia, topic-neutral. The outcome is a startlingly realistic three-dimensional arrangement where the viewer's change of perspective reveals more of the large cube as well as the small cube behind it.

But suppose that, having been told this story, our viewer consults his initial intuition (one which made the topic-neutrality of the cubes seem inconceivable) and decides, in the spirit of qualia realism, that the story does not 'satisfy' the intuition. For he can, after all, see that the virtual point source is there. He now declares himself a 'virtual point source realist'. At this point there is no further argument to undermine his 'virtual point source realism', but not because we grant that the intuition is to be employed as the ultimate measure of an explanation's success, rather, the illustration helps us realise that nothing can 'satisfy' or do full justice to the intuition except the intuition itself. This makes the intuition, if taken seriously, restrictive to such an extent that it halts all progress on the matter. Our viewer's 'virtual point source realism' would of course be ridiculed by scientists, but in more speculative fields such as philosophy of mind', appeals to similar intuitions tend to be much more charitably received. I do not deny, however, that they at least have some diagnostic value.

My aim has been to show that the qualia realists' continued reservations about qualitative content should be likened to the viewer's intuition in the above example. I of course make no claim to have established beyond doubt the nature of phenomenal properties in the same way that the laws of optics explain the nature of the cubes reproduced in a holographic image. The idea of qualitative content may turn out to be flawed, but I hope it is obvious enough that if this does happen, it is not the qualia realists' intuitions that will uncover the flaws.

I conclude with an observation that holography, setting aside now the illustration for the sake of which I began talking about it, perhaps hints at one useful way to think of the (non-qualitative) homogeneity of a first-order intentional object – whatever it is that causally mediates between a first-order state and the global meta-representation, instead of carrying information about the object as being divisible into numerous point sources on the 'surface' of the intentional object (i.e. as having a microstructure), the causal mediation only carries information about a single (topic-neutral) 'macro-source', namely, the unstructured homogeneous intentional object itself (in holography, a point source is the smallest point on the surface of the object being recorded that is able to reflect light, such as a molecule of a certain size)

3.2 Reviewing the Circular Argument

Having laid out my alternative HOR position, I will now return to the circular argument and explain how the position, when fed through that argument (by replacing (ii) with the alternative I put forward), escapes the consequence stated in (iv) that closes the circle. The argument consisted of the following four claims:

- (i) Qualia are not intrinsic properties of perceptual states, and if so, they must be wide contents.
- (ii) The thesis that qualia are wide contents can be plausibly held only as a claim of higher- rather than first-order intentionalism.

(iii) To have a higher-order thought about one's perceptual state is to token an internalised public-language sentence whose meaning is holistic and normative.

(iv) Since higher-order perceptual thoughts are internalised public-language sentences, qualia are intrinsic properties of experiences.

I was keen to show in Chapter 2. that the form of HOR (as referred to in (ii)) that stretches from wide first-order qualitative contents to introspective thoughts implodes into phenomenal individuals and hence returns us to the beginning of Chapter 1. where Lycan's adverbial analysis is invoked to dissolve those very same phenomenal individuals. In the remainder of Chapter 1. I gradually proceeded towards claim (i).

We saw that in view of Sellars' thought about introspection, the form of HOR in question is only as defensible as the existence of phenomenal concepts. Firstly, in order to tie phenomenal first-order states closely enough to the (experiencing) self, the relevant self-reflexive states must contain concepts whose extension excludes non-qualitative perceptions. Secondly, FOR was abandoned in favour of HOR because it did not distinguish blindsighters as well as other non-conscious perceivers, which can be seen as cases of absent qualia tailored for testing these accounts, from normally sighted perceivers. HOR offered to find a difference by mentioning higher-order states in its characterisation of qualia, for which predicates excluding non-qualitative perceptions from their extensions are necessary. Besides, the extension of every such predicate must be stable across worlds (actual and possible) to pre-empt qualia inversions. And finally, only phenomenal concepts (purely recognitional and innate Mentalese concepts) have possession-conditions (Peacocke, 1992) that one can meet prior to and independently of learning a public language. Their absence would make phenomenal consciousness the sole preserve of language-using creatures.

I need not repeat that Sellars' non-mentalistic treatment of introspection as internalised public-language sentences brings the epistemological soundness of phenomenal concepts into question. He would hardly feel uneasy about the loss of the rigid type of designation characteristic of phenomenal concepts on which HOR in its present form relies, as he would dismiss qualia inversions as a pseudo-problem, a regrettable episode of the epistemologist's fiction that nurtures the Myth of the Given.

For the meaning of introspective predicates applied to experiences is exhausted in what is communicated in overt speech, leaving no ineffable residue.

Unless we could find an alternative, the absence of phenomenal concepts would deprive us of a reductive explanation, in functional/representational terms, of what many feel must be a natural phenomenon - experience belongs to the logical space of nature, as opposed to the logical space of reasons, to which introspection, as Sellars treats it, as well as thoughts in general belong (the understanding of the contrast between experiences and thoughts as one between the logical space of nature and the logical space of reasons is from McDowell (1994)).

Suppose now that qualia are higher- rather than first-order contents in the way specified in the previous section and that (ii) is restated accordingly. The central idea is that a property one is non-conceptually aware of in experiencing is qualitative if the way he as the subject is related to it is identical to the relation between the global metarepresentation and its content. It is qualitative because the subject's perspective towards the property amounts to the metarepresentation's pointing towards its content. Only if first-order perceptions are taken to be qualia-bearing within a HOR picture do we have to link qualia to concepts that meet the above conditions. Shifting qualia to the higher-order level allows me to give a HOR account independently of phenomenal concepts and introspection in general. None of the components out of which my account is constructed are conceptual, since the qualia-bearing higher-order state must be non-epistemic.

Although it is higher-order, it would be unreasonable to say that the idea of a global metarepresentation may be extended to account for the introspective faculty dedicated to experiences given that the metarepresentation is itself a qualia-bearing state. There is of course a sense in which it is correct to say that, on my account, sensory qualia are self-reflexive – they are the contents of a state that is aimed at other mental states (i.e. lower-order perceptions) rather than at external objects.

I reserve the term 'introspection' for the type of *conceptual* exercise which involves an *explicit reference* to one's own inner episode, thereby presupposing a folk-psychological theory of mind. So the predicate 'red' is introspective if its user applies it to an aspect of what, based on his theory of mind, he is non-inferentially aware of as an

inner sensory occurrence. But I also spoke of phenomenal concepts as though they were introspective despite the fact that their acquisition does not require that the user views himself as having inner episodes (see the characterisation of purely recognitional and innate Mentalese concepts). They are introspective even though, due to a lack of a theory of mind, the user never employs them in thoughts which contain an explicit reference to such episodes. If phenomenal concepts were possible, they would indeed count as introspective while flouting the above ‘explicit reference’ requirement, as every phenomenal concept rigidly designates a phenomenal property, i.e. something necessarily being an aspect of an inner episode, not of something non-mental such as an external object. For example, a phenomenal concept formed by means of brute recognition rigidly designates a certain phenomenal property even though the user lacks concepts referring to his inner occurrences, and can only attribute that property to worldly objects.

Sellars’ non-mentalist treatment traces the origin of introspective predicates to the intersubjective framework of physical objects and properties expressed by public-language concepts (‘looks-F’ is parasitic on ‘is-red’ – to say that x looks F is to take back the endorsement of ‘x is F’). Public-language colour terms do not designate rigidly. Things I overtly judge to be green differ in their perceived colour from those which my counterparts at various possible worlds, speaking the same language, overtly judge to be green. Possible-world variations of this kind mean that my public-language concept ‘green’ does not capture a phenomenal property (at least not ‘non-comparatively’ (see Chisholm, 1957)). Nor does it, for that reason, capture an aspect of an inner occurrence. The introspective predicates derived by Sellars from these ordinary colour terms are similarly non-rigid and it would be hard to see how they are deployed inwardly if internalised public-language sentences did not contain an explicit reference to an inner experiential episode to serve as the subject of predication. Introspective concepts lose their prominent role in my higher-order alternative. I can therefore accept that the introspective capacity is the outcome of internalisation of public-language sentences. A story of how we ordinarily come to understand ourselves as creatures with experiences can at best supplement, but certainly not undermine, my account of qualia.

What does the difference between cases of perceiving without experiencing and actual experiences consist in if we assume that my position is correct? On the version of HOR

my proposal replaces, qualia are properties of first-order states and their characterisation must connect them in the required ways with higher-order states. In the case of blindsight, the connection with the requisite higher-order states is missing. My claim, on the other hand, is that blindsighters' all first-order states are as intact as those of normally sighted humans. Even the latter's perceptions lack qualia. We are spared the task of finding such elements of higher-order states which reliably indicate, by means of rigid designation, the presence of qualia if we suppose that the meta-representation itself is qualitative. Being blindsighted means that a certain portion of the meta-representation has been lost, perhaps due to the links between perceptions and the metarepresentation having been partially severed.

If claim (iii) (stating that to have a higher-order thought about one's perceptual state is to token an internalised public-language sentence whose meaning is holistic and normative) is true of introspection but is irrelevant to experience, (iv) will not follow. My proposal enables me to capture in (higher-order) relational terms what non-conscious perceivers lack in comparison with sighted creatures, avoiding intrinsic qualia. What is more, since the difference is so captured while staying within the framework of states rather than objects, it is clear that the above reintroduction of phenomenal individuals (see 2.3 *Sensa and Phenomenal Individuals*) is no longer enforced on us. It is no longer inevitable that the difference lies in the fact that whereas we have sense data, non-conscious perceivers don't.

3.3 Higher-Order Contents Without Higher-Order States

The position I put forward in the previous section had been shaped primarily by Sellars' worries about the existence of phenomenal concepts as well as the existence of the self serving as the subject of experience. In advancing my alternative version of higher-order representationalism, I avoided relying on phenomenal concepts by suggesting that the higher-order structure of phenomenal consciousness is non-conceptual throughout, while treating the introspective capacity as a separate (and essentially conceptual) matter not involved in that structure. We found sufficient evidence in Sellars'

work that he would dismiss phenomenal concepts, which are delivered ‘privately’ and independently of any inferential links with other concepts, as an instance of placing something allegedly epistemic outside the logical Space of Reasons. Since the states that I do take to belong in the HOR structure of phenomenal consciousness are non-conceptual, I am not forcing anything epistemic outside the Space of Reasons.

Of course, my claim that the higher-order states in question are non-epistemic is not in itself a novelty. Lycan’s (1996) is also a presentation of a higher-order perception model. I have already noted, however, that it does not overcome the problem of the self to which we attribute experiences. He has it that higher-order states are the outcome of the operation of internal monitors (or Inner Sense). The monitors as well as everything they interact with are sub-personal parts rather than the sort of persons to which experiences are attributable.²⁶ Recall from Chapter 2. how Dennett (1978) stated some such difficulty:

At best a sub-personal theory will seem to give us *no grounds* for believing its instantiations would be subjects of experience, and at worst (as we have seen) a sub-personal theory will seem to permit instantiations that *obviously are not* subjects of experience (p.154, italics original).

Nor is the *content* of the monitors’ outputs a personal-level one – they are intended to explain introspective attention in Lycan’s theory, and the amount of perceptual and proprioceptive intake introspection can focus on at a given time is therefore limited, in other words, the monitors’ introspective content is local rather than global at any given time. This is where my emphasis on global higher-order content comes into play. I observed above that once we think of the higher-order level as a global meta-representation, it is open to us to say that although the meta-representation is sub-personal (as it is a brain state), its content is *personal* due to the fact that it is all-encompassing (i.e. it covers one’s entire perceptual and proprioceptive output) and

²⁶ The terms ‘personal’ and ‘sub-personal’ are often used to distinguish, say, early visual processing from the output of visual processing. The former has no direct psychological significance for the subject, whereas the latter become a part of the subject’s mental life. Clearly, it is not this distinction that I have in mind in the present context.

higher-order.²⁷ The main strength of this answer is that it replaces the arguably insurmountable problems regarding the self to which we ordinarily attribute experiences with far more tractable questions about global meta-content. It also squares rather nicely with Sellars' claim that, since we consist of a multitude of micro-physical parts, there is no unified self serving as the subject of experience. Positing a global state does not contradict that claim. Nor does my answer rely on building the self out of the inevitably sub-personal neural counterparts of the components of a computing machine (such as a mind-reading module or a system of internal monitors).

Unlike Lycan's internal monitors, the global meta-representation is not meant to function as an inner attention mechanism – it is not to explain the difference between states we are aware of being in and those we aren't. It is continuously aimed at all perceptual and proprioceptive outputs and I also take it to be qualia-bearing. Introspection, on the other hand, is an altogether distinct mechanism. As I understand it, following Sellars, it is purely conceptual, requiring that one be a member of a linguistic community.

Another important point made in the preceding section concerns my shifting qualia to the higher-order level. I said that a position on which there is a non-qualitative (non-epistemic) global meta-representation alongside qualia-bearing first-order states would be unworkable. For the sake of illustration, think of the global meta-content as being superimposed over all lower-order contents. Now, if it was a non-qualitative meta-content that was so superimposed over qualitative lower-order content, the outcome would be a quite unfortunate situation where we suffer from a condition that is the reverse of blindsight, namely, 'sightblind', as I called it above. Because lower-order contents would be continuously covered by a global higher-order 'layer' of content that lacks qualia, we would have phenomenal states at the lower level and yet not experience anything. Similarly, allowing the presence of qualia at both levels would mean that lower-order qualia would not be a part of our inner lives.

Lycan's (1996) has been criticised (see (Neander 1998)) on the grounds that the Inner Sense – the inner sensing being representational rather than non-representational – should

²⁷ An objection may be raised here that a thoroughly global content should also cover thoughts, beliefs, etc., but it is a characterization of the experiencing self that I'm interested in, one which I ought to be able to provide regardless of conceptualization.

produce its own qualia alongside lower-order ones. If the introspective mechanism misrepresents and there is, for example, a green lower-order quale and a red higher-order quale, it becomes unclear what exactly the perceiver is experiencing. This, however, is unfair, for Lycan does stress that internal scanning produces representations in the Language of Thought. So the contents of these representations must be understood as epistemic and hence by no means as qualitative. Introspection is perception-like because the way it shifts from one sensation to another resembles the way the eyes shift focus from one part of a scene to another, not because it produces non-conceptual representations. It is worth reminding that only if lower-order qualia were sensed non-representationally – much like sense-data – could we have non-epistemic introspection that lacks qualia. But of course Lycan's programme, as he announces it, is that of establishing the hegemony of representation, not that of establishing a 'mongrel' position with elements of representationalism (in the case of perception) and elements the sense-datum theory (in the case of introspection).

With these summarising remarks in mind, I shall go on to consider some notable overlaps with the work of Robert Van Gulick and Peter Carruthers. There are also equally notable disagreements to be recognised regarding the way they would approach the challenge my position is intended to deal with. Common to their proposals is the idea of higher-order qualitative contents without higher-order states. I will explain a little later why I reject the idea.

Van Gulick (2000, 2004) is the only author who mentions global states, although he is interested in sketching a model of introspective awareness not in a theory of qualia. As we will appreciate shortly, it classes neither as a HOT model (higher-order thought) nor as a HOP model (higher-order perception). It draws heavily on Dennett's (1991) multiple drafts theory. The key assumption is that all qualia, whether introspected or not, contain a subjective element in that they presuppose the existence of a subject (in my account the subject as it is traditionally understood is replaced with a global higher-order state). I have of course said on many occasions that this is the essential characteristic of qualia, both heeded and unheeded. Van Gulick calls it a 'self-referential element', but it is clear from his writing that it is subjectivity in the above sense that he speaks of.

The most striking difference between standard higher-order models and Van Gulick's thought is that according to him higher-order contents occur in the absence of actual higher-order states. How can a first-order system generate higher-order contents? Suppose there is a (first-order) phenomenal visual state *S* implicit in which is the aforementioned subjective element. *S*'s causal potential is initially limited; it has, as it were, a local causal impact on a relatively small number of states. However, *S* can be integrated into a larger causal network of states where its causal potential increases. In particular, the kind of circumstances that make integration possible are those in which a state becomes a part of one's current flow of consciousness. One's flow of consciousness at a given moment is a set of states which together form what Van Gulick calls a 'global state'. Even though the individual content-bearing states in one's current flow of consciousness are first-order, they have a global causal potential by virtue of being parts of that flow. As their sphere of causal influence increases, they gain a more central role in the system. In the case of a phenomenal state like *S*, gaining a more central role means that the subjective element implicit in its quale is accentuated, so to speak, to such an extent that the content is transformed into a higher-order one – it is so transformed even without the occurrence of a higher-order state. Since this is presented as a model of introspection, the first-order states that together form one's flow of consciousness (i.e. one's global state) have introspective contents. So Van Gulick's global state is somewhat less global than the meta-representation as I described it when making my own proposal. I introduced it as representation that is continuously aimed at all sensory and proprioceptive stimulation, whether introspectively attended or not, and held that introspection is a conceptual capacity independent of that meta-representation.

As for the transformation of first-order contents into higher-order ones, Van Gulick says the following:

Nonetheless the HOGS model remains a type of higher-order theory in so far as the change that occurs in a lower-order state's function as it is integrated into the momentary global correlate of self-awareness transforms its content in ways that involve a heightened element of experiential self-reference (Van Gulick, 2000).

To some degree this self-referential aspect may be present in a limited and wholly implicitly way even in states that we typically regard as nonconscious. Even nonconscious perceptions or a nonconscious desires may incorporate some implicit elements of self-reference in their satisfaction conditions: a complex such as “me-seeing: a tree/ here/now” or “me-desiring: a drink of water/here/now” may do a better job of capturing the intentionality of such nonconscious states than would the merely objectual “tree/here/now”. In so far as this is so, the change in content that accompanies the move from unconscious to conscious status is not the totally *de novo* addition of a self-reflexive aspect but rather the transformation from a limited and implicit self-referential aspect to a richer and more explicit one (2000).

It follows then that both (lower-order) perceptual contents and the relevant introspective contents are qualitative, the difference between the respective qualia lying in the degree of explicitness of the subjective aspect, not in the occurrence of a discrete higher-order state. But a lower- and a higher-order quale do not occur simultaneously in a given experiential situation, for when a lower-order perceptual state joins a larger causal network (of first-order states), its quale is transferred to the higher-order level (although of course the state itself does not undergo a similar shift).

Because they do not occur simultaneously, Van Gulick avoids the criticism of HOP models I mentioned in connection with Lycan’s internal monitoring. The problem was that if a red quale is misrepresented by a non-epistemic introspective state as a green quale, it is unclear which of the two determines the character of my present experience. In the case of Van Gulick’s model, for every experiential situation there is only one qualitative content which is either lower- or higher-order depending on how global or local the causal network it belong in is, hence no such problem arises here.

I will now argue that the objections that constituted my challenge for representationalism reapply to Van Gulick’s model. I have already said that I agree with his claim that all qualia, regardless of whether they are the objects of introspective attention or not, contain a subjective element to the effect that they presuppose the

existence of a subject to which sensations are attributed. In the quoted passage the subjective element is expressed as “me-seeing: a tree/ here/now”, and this seems to me to be unproblematic. Now, Van Gulick holds that higher-order qualia differ from lower-order ones in that the subjective element is present in them in a more explicit manner, in other words, it is something about the quale itself, something which is ‘intrinsic’ to it in some sense, that makes it higher-order, not something about the qualia-bearing state; there is no corresponding adjustment in the qualia-bearing state – it (as well as the global causal network in which it belongs) remains first-order. Thus we should expect the more explicit presence of the subjective element, which qualia have ‘intrinsically’, to show up in a certain distinctive way in the character of my experience. I nonetheless don’t see how it does.

Suppose that, looking at a tree, I am in a lower-order state with a green quale to which I’m not paying introspective attention. The phenomenal greenness is experienced as a property of the tree I’m looking at. Let us state the subjective element present in the phenomenal property in the same way as Van Gulick does, i.e. “me-seeing: a tree/here/now”. Suppose also that I turn my attention inwards and that the changes taking place in my brain as I do so can be exhaustively described using Van Gulick’s model and, as a result, the green quale turns into a higher-order content (without the occurrence of a higher-order state). Van Gulick would say that it is not the case that I have entered into a new state, a higher-order one, all that has changed is that my quale now has a strengthened subjective aspect. But just how does this change translate into how the phenomenal property is experienced given that I continue to experience it as a property of the tree? Having turned my attention inwards, I simply fail to notice a more ‘tangible’ subjective element which contrasts my quale with the lower-order quale my state featured before I started introspecting.

The fact that it is impossible to distinguish higher-order qualia from lower-order ones in terms of the degree of explicitness of the subjective aspect alone rather than in terms of the occurrence of a higher-order qualia-bearing states leads to the blindsight-based objection to first-order representationalism raised, following Carruthers, in Chapter 2. Blindsighters do not strictly satisfy the set of conditions that the acronym ‘PANIC’ (Tye 1995) stands for, more specifically, they do not satisfy the requirement that phenomenal

perceptual states have an impact on first-order beliefs and actions, since they do not form beliefs with respect to stimulation in the blind portion of their visual field. In spite of this, the concession appeared to be a minor one, as they do show substantial behavioural sensitivity to external stimulation in view of which first-order belief-formation can hardly be a decisive factor in attributing or not attributing wide contents to their 'blind' states. Since their behavioural sensitivity suffices to warrant attribution of wide contents, their states should class as phenomenal even though they apparently aren't. This meant that the locus of phenomenal consciousness, as Carruthers (2000) says, is experiential subjectivity not worldly subjectivity and that we need to formulate a higher-order condition to distinguish ourselves from blindsighters. The condition with which I began my examination of HOR was that a perceptual state is qualitative if it is available to be picked out by a higher-order representation, although I realised in the discussion that followed that not all types of higher-order representations apply distinctively to qualitative (as opposed non-qualitative) perceptions.

An argument along some such lines also undermines Van Gulick's position despite having been put forward as higher-order representationalism. The first obvious difficulty is that he only speaks of higher-order contents while higher-order states are absent. He might argue that his view does not imply that blindsighters' states are phenomenal because they are not available (or poised we may say) to be integrated into a global representation, they do not join one's flow of consciousness, as their causal impact is limited at best and remains such. But again their behavioural sensitivity is sufficient to justify attribution of content and the states' not having a global impact on other first-order states is unlikely to be an overriding reason not to attribute content. Another move Van Gulick might make is to say that the source of the required higher-order condition is this: When an individual perceptual representation joins a global representation, its content is transformed into a higher-order one, which does capture experiential subjectivity even in absence of higher-order representations. Blindsighters clearly do not satisfy this condition.

I must point out, in response to this move, that the only criterion by which Van Gulick differentiates between higher- and lower-order qualia is the degree of explicitness of the subjective element. However, I struggled to see how we can specify any distinctive

way the explicit subjective element alters the character of my experience – phenomenal colours are still experienced as features of worldly objects; and I noted that it is not open to him to say that what sets higher-order qualia apart from first-order qualia is that they are contents of higher-order representations, for such representations are missing. Consequently, if there is nothing to distinguish his position from FOR – the difference does not lie in there being higher-order representations, nor does it lie in the character of ‘higher-order’ qualia – the blindsight-based objection is reinstated.

On the other hand, my account does not imply that higher-order qualia differ ‘intrinsically’ (in the degree of explicitness of the subjective element) from qualia construed as first-order contents – both contain the subjective element in equal measure. Global meta-representational qualia are experienced as features of worldly objects and what sets them apart as higher-order is the presence of an actual meta-representation. But I also deny that first-order perceptual contents are phenomenal.

In view of Sellars’ reconstruction of our introspective conceptions regarding experiential episodes, there is another (rather major) flaw in Van Gulick’s position to draw attention to. Recall that Van Gulick’s primary interest is in giving an account of introspective awareness with respect to experiences, so the global higher-order content he speaks of is in fact introspective on his view. Since such an introspective content is qualitative and hence necessarily non-conceptual, it does not require that we have a theory of mind in order to be able to introspect. To have a theory of mind is to have a certain understanding (which of course presupposes conceptualisation) of inner experiential episodes. It is by virtue of having a theory of mind that we see ourselves as creatures who have feelings and sensations and, crucially, are able to turn our attention inwards. So Van Gulick’s claim that introspection is non-epistemic inevitably raises the question of what exactly it is that guides my attention inwards if I have no understanding of anything ‘inner’, let alone of any inner episodes. To this he can give no satisfactory answer²⁸. Van Gulick does (in his 1988) offer a general functionalist account of

²⁸ I should explain that Van Gulick uses the term ‘introspection’ in a much narrower sense than I have been doing throughout my work. He understands introspection as just one form of self-awareness, one which consists in actively focusing inner attention on one’s own thoughts and experiences. More passive cases of self-awareness where one’s focusing his attention inwards is prompted by, say, an intense pain would not class as introspective. So it is self-awareness in general that the proposal under discussion is to account for. On the other hand, ‘introspection’, as I have been using the term (throughout as well as in this

introspection elsewhere, one which is not tailored for, and does not imply, for that matter, any specific theory of qualia. In that paper introspection is not conceived of as a non-epistemic faculty, so it could not be so obviously criticised on the same grounds as the model under discussion. Unfortunately, the general account by no means complements the present model and hence does not vindicate it.

For these reasons, I was wary of entangling the global meta-representation with the introspective capacity when presenting my account. The global meta-representation, on my account, is innate, which means that my higher-order account is such that it has no astonishing implications for phenomenal consciousness in small children and some animals, while our introspective capacity develops later in life with the acquisition of public-language.

The problem of illegitimately forcing a conceptual faculty (i.e. introspection) outside what Sellars calls the logical space of reasons does not beset Carruthers' (2000) higher-order model. His model is another instance of higher-order qualia without higher-order states. His story of how higher-order contents are generated in absence of higher-order states differs from Van Gulick's in such a way that it avoids mixing non-epistemic phenomena with epistemic ones. To begin, Carruthers holds that lower-order perceptual contents are non-qualitative, which is one important overlap with my proposal. Now, a pre-linguistic child's lower-order perceptual states have non-qualitative analogue contents such as 'red_a' (where the index 'a' stands for analogue). When a child has learnt his first language, he eventually comes to have a grasp of the is/seems distinction resulting in the development of a theory of mind, whereby he acquires the ability to conceptualise and refer, both covertly (in thought) and overtly (in verbal reports), to his experiences. This ability turns all of his non-qualitative perceptual contents (such as 'red_a') into *qualitative* higher-order ones (such as 'experience of red_a'). It is not that 'experience of red_a' is the content of the state which constitutes the child's referring to his experience (it is, after all, out of the question that epistemic states could have non-epistemic contents), rather, the thought is that when the lower-order perception with

section), is interchangeable with 'self-awareness', so passive cases of self-awareness also qualify as introspective. But nothing important depends on this terminological difference, for my criticism is directed at Van Gulick's treatment of passive forms of self-awareness. Even the sharpest pain won't turn my attention inwards if I have no understanding of anything 'inner'.

content 'red_a' causally connects with the newly-formed theory-of-mind module, this new causal/functional connection alters its content to such an extent that it becomes a higher-order one. If we suppose, as all functionalists do, that content is determined by functional role, and that an adjustment in a state's functional role brought about by a new causal/functional link also alters the state's content, the claim is not at all implausible.

We notice at once that unlike my proposal, Carruthers' model does have startling implications for phenomenal consciousness in pre-linguistic children and some animals. Small children perceive without experiencing, much like blindsighters, until they have a full-fledged theory-of-mind module. Secondly, it does not address the issue of identifying the subject of experience, since the theory-of-mind module and other module it interacts with are sub-personal parts. I tried to address it with the notion of a global higher-order content.

Some may be bewildered by Carruthers' suggestion that infants develop their theory of mind with respect to perceptions which aren't yet experiences. For it is only when the theory-of-mind module is in place that their lower-order non-qualitative contents are transformed into higher-order qualitative ones. So the process of the formation of the module is underway while they still don't experience anything. Nonetheless, Sellars' story of how we come to have the sort ordinary theory of mind on which our introspective conceptions are based offers a response, although Carruthers doesn't refer to his work. Central to the story was the idea that inner experiential episodes are conceptualised as unobserved theoretical posits which later become non-inferentially reportable, just as someone working in a laboratory may be disposed to report the presence of NaCl rather than salt. Stephen Leeds' main contribution in (1993), as we saw in 2.2.3 (*HOR's Commitment to Recognition Concepts*), was the argument that nothing in Sellars' reconstruction precludes blindsighters from learning to non-inferentially report previously unobserved perceptual episodes in the blind portion of their visual field given that the reconstruction does not involve phenomenal concepts at any stage. Nor is there anything which precludes pre-linguistic children from developing a theory of mind with respect to perceptions which are not yet experiences.

Conclusion

Having reached the very end, let me repeat briefly what the main line of argument in my work has been. Chapter One was a way of showing how we can get from intrinsic qualia to qualia viewed as wide contents. Although no knockout argument is offered at any stage, all of the objections that lead to claims that qualia, or qualia as well as experiential contents, aren't wide can be overcome.

The beginning of Chapter Two is a suggestion that far more insight into the notion of qualitative content is to be gained if the debate in the current literature on representationalism shifts its focus to how representationalists can distinguish between non-qualitative perceptions and experiences. The failure to distinguish them in first-order representational terms means that we should begin by exploring whether qualitative perceptions connect with higher-order states in ways that non-qualitative ones don't. This strategy will prove fruitful, however, only if we can establish that it is higher-order states containing phenomenal concepts that qualitative perceptions connect with. I then argue that Sellars' thought regarding introspection must be taken into consideration in deciding whether phenomenal concepts can be accepted as epistemologically sound. We saw that the appeals to phenomenal concepts I discuss repeat the mistakes of classical empiricism, which rules them out as a possible distinguishing factor. Provided one finds each step of the argument so far convincing enough, it appears that first-order qualia connect with higher-order thoughts are supplied by public language, not by an act of brute recognition or innate Mentalese. But non-qualitative perceptions aren't precluded from forming connections with higher-order thoughts so conceived either, at which point the strategy of identifying a higher-order connection that is unique to qualitative ones seems to have reached its limit. Remember that Chapter One I argued that (wide content) representationalism is the most robust reductive strategy we have, and hence that if qualia are extrinsic properties of experiences, they must be wide contents. However, even the higher-order condition cannot sustain qualia taken as wide contents, in which case it must be intrinsic qualia that phenomenal states have and non-phenomenal perceptions don't.

I make my contribution in Chapter Three where I make my alternative representationalist proposal. The idea is that in order to avoid reintroducing intrinsic

qualia we should abandon the view that qualia are first-order contents. The first-order states of normally sighted humans lack qualia, and so they do not differ from the states of blindsighters in this respect. They differ from those of blindsighters in that their (wide) intentional contents are represented by a global higher-order state, the content of which global state is qualitative. So when wide contents are represented at the higher-order level, the meta-representational content is qualitative. In the case of blindsight, the causal link between some of their first-order states and their global meta-representation is broken.

Bibliographic References:

- Armstrong, D. M.: *Bodily Sensations*, London: Routledge and Kegan Paul, 1962
- Armstrong, D. M.: *What Is Consciousness*: IN: *The Nature of Mind*, Ithaca, NY, Cornell University Press, 1981.
- Armstrong, D. M.: *Consciousness and causality*. IN: D. Armstrong and N. Malcolm, *Consciousness and Causality*, Blackwell, 1984.
- Block, Ned: *A Confusion About a Function of Consciousness*, *Behavioral and Brain Sciences*, 1995.
- Block, Ned: *Inverted Earth*, IN: *Nous-Supplement: Philosophical Perspectives*, vol. 4, 1990.
- Block, N. and Stalnaker R.: *Conceptual Analysis, Dualism, and the Explanatory Gap*, IN: *The Philosophical Review*, 1999.
- Brandom, Robert: *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Cambridge, MA, Harvard University Press, 1998.
- Burge, Tyler: *Individualism and the Mental*, IN: *Midwest Studies in Philosophy*, 4, 73-121, 1979.
- Burge, Tyler: *Individualism and Psychology*, IN: *The Philosophical Review*, Vol. 95, No. 1, pp. 3-45, 1986.
- Chalmers, D.: *The Conscious Mind*, Oxford University Press, 1996.
- Chalmers, D.: *Perception and the Fall from Eden*, IN: T. Gendler & J. Hawthorne, eds, *Perceptual Experience*, Oxford University Press, 2006.
- Chisholm, Roderick: *Perceiving: A Philosophical Study*, Ithaca, New York: Cornell University Press, 1957.
- Carruthers, P.: *Language, Thought, and Consciousness: An Essay in Philosophical Psychology*, Cambridge, 1996.
- Carruthers, P.: *Phenomenal Consciousness: A Naturalistic Theory*, Cambridge, 2000.
- Carruthers, P.: *Phenomenal Concepts and Higher-Order Experiences*, IN: *Philosophy and Phenomenological Research*, 2004.
- Castaneda, Hector-Neri: *He: A Study in the Logic of Self-Consciousness*, *Ratio* 8: 130-157, 1966.
- Clark, Austen: *Sensory Qualities*, Oxford: Oxford University Press, 1993.
- .
- Davies, Martin: *Language, Thought, and the Language of Thought (Aunty's Own Argument Revisited)*, IN: P. Carruthers and J. Boucher (eds), *Language and Thought*, 1998 (Cambridge: Cambridge University Press, 1998), 226–47.
- Dennett, D.: *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge: Bradford Books/MIT Press, 1978.
- Dennett, D.: *Consciousness Explained*, Allen Lane, 1991.
- Devitt, Michael: *Meanings Just Ain't in the Head*, IN Boolos, G., ed., *Meaning and Method: Essays in Honor of Hilary Putnam*, Cambridge: Cambridge University Press, 1990.
- Dretske, Fred: *Naturalizing the Mind*, Cambridge (MA): MIT Press, 1995.

- Dulany, D. E.: Higher Order Representation in a Mentalistic Metatheory, IN: Gennaro, Rocco: Higher-Order Theories of Consciousness: An Anthology, John Benjamins Publishing Company, 2004.
- Fodor, Jerry: The Language of Thought, Cambridge, Massachusetts: Harvard University Press, 1975
- Fodor, Jerry: Psychosemantics: The Problem of Meaning in the Philosophy of Mind, MIT Press, 1987.
- Fodor, Jerry: Concepts: Where Cognitive Science Went Wrong, Oxford University Press, 1998.
- Gibbard, Allan: Meaning and Normativity, Philosophical Issues 5, Enrique Villanueva, ed., Truth and Rationality, Atascadero, CA: Ridgeview Publishing Co., 1994.
- Harman, Gilbert: Some Philosophical Issues in Cognitive science, IN: Foundations of Cognitive Science, ed. Michael I. Posner, Cambridge (MA): MIT Press, 1989.
- Jackson, Frank: Epiphenomenal Qualia, IN: Philosophical Quarterly, 1982.
- Jackson, Frank: Mental Causation, IN: Mind, 1996.
- Kripke, Saul: Naming and Necessity, Cambridge: Cambridge University Press, 1972.
- Leeds, Stephen: Qualia, Awareness, Sellars, IN: Nous, 1993.
- Levine, Joseph: Purple Haze: The Puzzle of Consciousness (New York: Oxford University Press, 2001)
- Levine, Joseph: On Leaving Out what It's Like, IN: Davies and Humphreys, 1993.
- Levine, Joseph: Conceivability and the Metaphysics of Mind, Nous, 1998.
- Loar, B.: Phenomenal States, IN: Tomberlin, 1990.
- Loar, B.: Phenomenal States (Second version), IN: Block et al, 1996.
- Lycan, W. G.: Consciousness, Cambridge: MIT Press, 1987.
- Lycan, W. G.: Consciousness and Experience, Bradford, MIT; Cambridge Mass, 1996.
- Lycan, W. G.: Consciousness as Internal Monitoring, IN: Ned block, Owen Flanagan, and Guven Guzeldere: The Nature of Consciousness Philosophical Debates MIT Press, 1997.
- Marcel, A.: Blindsight and Shape Perception: Deficit of Visual Consciousness Or of Visual Function? IN: Brain, 121, 1998.
- Marr, D.: Vision, Henry Holt & Company, 1983.
- McDowell, John: Mind and World, Cambridge, Mass.: Harvard University Press, 1994.
- McLaughlin, M. and Tye, M.: Externalism, Twin Earth, and Self-knowledge, IN: Knowing Our Own Minds, ed. C. Wright, B. Smith and C. Macdonald, 285–320, 1998.
- Milner, D and Goodale, M.: The Visual Brain in Action, Oxford University Press, 1995.
- Neander, Karen: The Division of Phenomenal Labor, IN: Philosophical Perspectives, 12, Language, Mind , and Ontology, 1998.
- Perry, John: Cognitive Significance and New Theories of Reference, *Noûs* 2, no. 2, pp. 1 – 18, 1988.
- Papineau, D.: Philosophical Naturalism, Basil Blackwell, 1993a.
- Papineau, D.: Physicalism, Consciousness and the Antipathetic Fallacy, Australian Journal of Philosophy, 1993b.
- Pitcher, George: Pain Perception, IN: Philosophical Review 79, 1970.
- Peacocke, Christopher: Sense and Content, Oxford: Oxford University Press, 1983.
- Putnam, Hilary: Minds and machines, IN: Sidney Hook (ed.), Dimensions of Mind, New York University Press, 1960.

- Putnam, H.: Psychological Predicates, IN: Art, Mind and Religion, ed. W. H. Capitan and D. D. Merrill, Pittsburgh, Penn.: University of Pittsburgh Press, 1967.
- Putnam, H.: The Meaning of Meaning, IN: Philosophical Papers, vol. 2, Cambridge: Cambridge University Press, 1975.
- Putnam, H.: Reason, Truth and History, Cambridge: Cambridge University Press, 1981.
- Putnam, H.: Representation and Reality, MIT Press, 1991.
- Rey, Georges: A Narrow Representationalist Account of Qualitative Experience, IN: Philosophical Perspectives, 12, Language, Mind , and Ontology, 1998.
- Rey, Georges: Contemporary Philosophy of Mind: A Contentiously Classical Approach, Oxford: Blackwell, 1997.
- Rosenthal, D.: Two Concepts of Consciousness, IN: Philosophical Studies, 49, 1986.
- Rosenthal, D.: Consciousness and Mind, Clarendon Press, 2005.
- Searle, John: Intentionality: An Essay in The Philosophy of Mind, Cambridge: Cambridge University Press, 1983.
- Sellars, W.: A Semantical Solution of the Mind-Body Problem”, Methodos 5, pp. 45–82, 1953.
- Sellars, W.: Phenomenalism, IN: Science, Perception and Reality, London: Routledge and Kegan Paul, 1963.
- Sellars, W.: Language as Thought and as Communication, IN: Philosophy and Phenomenological Research, 1969.
- Sellars, W.: Empiricism and the Philosophy of Mind, Harvard University Press, 1997.
- Shewmon, D.: Consciousness in Congenitally Decorticate Children, IN: Developmental Medicine and Child Neurology, 41, 364-374.
- Shoemaker, Sydney: The Inverted Spectrum, IN: The journal of Philosophy, vol. 79, No. 7, 1982.
- Shoemaker, Sydney: Qualities and Qualia: What’s in the Mind? IN: Philosophy and Phenomenological Research, 50, Supplement, 1990.
- Sturgeon, Scott: The Epistemic View of Subjectivity, IN: The Journal of Philosophy, vol. 91, No. 5, 1994.
- Sturgeon, Scott: Matters of Mind: Consciousness, Reason and Nature, Routledge, 2000.
- Tye Michael: The Imagery Debate, Cambridge: MIT Press, 1991.
- Tye, Michael: Visual Qualia and Visual Content, IN: The Contents of Experience, edited by Tim Crane, 1992.
- Tye, Michael: Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind, Cambridge: MIT Press, 1995.
- Tye, Michael: The Problem of Simple Minds: Is There Anything It Is Like To Be A Honey Bee? In: Philosophical Studies, 88, 289-317, 1997
- Tye, Michael: Inverted Earth, Swampman and Representationism, IN: Philosophical Perspectives, 12, Language, Mind , and Ontology, 1998a.
- Tye, Michael: Externalism and Memory, IN: Proceedings of the Aristotelian Society, Supplementary Volume, 1998b.
- Tye, Michael: Phenomenal Consciousness, IN: Mind, 1999.
- Van Gulick, Robert: A functionalist plea for self-consciousness, IN: Philosophical Review 97: 149-88, 1988.
- Van Gulick, Robert: Inward and Upward: Reflection, Introspection, and Self-Awareness, Philosophical Topics 28:275-305, 2000.

- Robert Van Gulick: Higher-Order Global States (Hogs): An Alternative Higher-Order Model of Consciousness, IN: Rocco J. Gennaro (ed.), Higher-Order Theories of Consciousness: An Anthology, John Benjamins, 2004.
- Weiskrantz, L.: Blindsight, Oxford University Press, 1986.
- White, Stephen: Color and Notional Content, IN: Philosophical Topics, vol. 22, No. 1&2, 1994.

