# CONSCIOUSNESS, PERCEPTION, AND SHORT-TERM MEMORY

**Henry Shevlin**

Dissertation submitted for the degree of Doctor of Philosophy in the Department of Philosophy, CUNY Graduate Center, 2016.

*For Wanda and my parents, without whom this dissertation would not have been written.*

*And for Luke, without whom it would have been written a lot sooner.*

# TABLE OF CONTENTS

# INTRODUCTION

The notion of memory, as it is used in ordinary language, may seem to have little directly to do with perceptual experience. While perceptual experience informs us about the world as it is now, memory almost by definition tells us about the past. Similarly, whereas perceptual experience seems like an occurrent phenomenon, existing fleetingly from one moment to the next, it's natural to think of memory as being more like an inert store of information, accessible when we need it but capable of lying dormant for years at a time.

This way of thinking about experience and memory certainly tracks typical uses of the word "memory" in ordinary language. However, there is a broader notion of memory that includes both long-term processes of storage and the short-term retention of information. William James (1890) made this observation, distinguishing between "primary memory" – the fleeting, momentary impressions that make up the "specious present" – and "secondary memory", the store of past impressions from which we can recall past experiences or acquired information. Contrasting the two, he states that "[w]ithin the few seconds which constitute the specious present there is an intuitive perception of the successive moments. But these moments, of which we have a primary memory-image, are not properly recalled from the past, our knowledge of them is in no way analogous to a memory properly so called."

Similarly, in contemporary cognitive science, memory is taken to include almost any psychological process that functions to store or maintain information, even if only for very brief durations. In this broader sense of the term, the connections between memory and conscious experience should be more apparent. After all, some mechanism for the short-term retention of information will be required for almost any perceptual or cognitive process such as recognition or inference to take place: as one group of psychologists put it, "storage, in the sense of internal representation, is a prerequisite for processing" (Halford, Phillips, & Wilson, 2001). Assuming, then, as many theorists do, that consciousness consists at least partly in the occurrence of psychological processes, memory is likely to have an important role to play in a scientific theory of consciousness.

Moreover, on closer examination, the claim that short-term memory is intimately connected to perceptual experience may seem to have some intuitive and phenomenological appeal. When looking over a selection of pastries to decide which looks the most appetizing, I seem to hold them simultaneously in mind as I compare them; when I decide how to respond to a friend's question, her words seem to linger in my consciousness for a second. And even just watching a sunrise, my perceptual experience is briefly poised to evoke memories or reflections.

The goal of this dissertation will be to explore the connections between consciousness, perception, and short-term memory, and specifically to argue that a controversial new form of short-term memory may provide us with a novel window into conscious experience. I begin, in Chapter 1 by summarizing some of the key findings of the voluminous but scattered research on memory, condensing them into a broad functional architecture that I term the Sensory-Cognitive Model. I then briefly show how this model can inform philosophical projects relating to the perception-cognition border and consciousness. In Chapter 2, I examine recent psychological evidence that raises the possibility that we should expand the Sensory-Cognitive model to include a further stage of processing termed Conceptual Short-Term Memory (CSTM). In Chapter 3, I suggest ways in which CSTM may help us understand a puzzling form of experience at the cognition and perception boundary, arguing specifically for the claim that CSTM may underlie what is called *categorical perception* or 'perceiving-as'. I consider how this way of thinking about categorical perception may relate to discussion of the controversial claim that perceptual experience includes high-level phenomenal properties, and how it might inform debates about the cognitive penetration of experience. In Chapter 4, I turn to the theories of consciousness debate. After examining the leading theories, I suggest how CSTM may point to novel approaches, and spell out a new theory of consciousness I call the Workspace-Plus account. Finally, in Chapter 5, I turn to broader issues concerning the place of CSTM in nature. After showing how my account can be spelled out in terms of the framework offered by Tyler Burge, I examine what a CSTM view of consciousness may have to say about consciousness in animals. I close with an investigation of how this view of consciousness might be applied to a particular problem of animal minds, namely the evaluation of suffering in non-human animals.

In proceeding, it is worth briefly spelling out the scope of this dissertation. As noted, the unifying theme of this project is to argue for a new form of short-term memory and spell out how it can be applied to theorizing about the mind. This will involve discussing a wide range of experimental and theoretical literature. This project is only tractable, however, because there are certain questions that I have chosen not to directly consider. Two such questions stand out.

First, perhaps most importantly, I do not attempt to give a theory of concepts or a fully worked-out account of the relationship between conceptual and nonconceptual content. This may seem a striking omission, given that the protagonist in the theoretical story I will telling is *Conceptual* Short-Term Memory. However, attempting to give a proper account of what it is that makes a given representation conceptual would too complex and fraught an endeavor to take on in this dissertation given my other goals. Still, I will not be able to escape these questions entirely: I will frequently have recourse to these terms insofar as they are used by various theorists to describe their own accounts, and Chapters 4 and 5 in particular include some brief reflections concerning how the conceptual/non-conceptual distinction may be relevant to my project. Nonetheless, for the most part, especially early in the dissertation, whenever I use the words ‘concept’ or ‘conceptual’, I will be deferring relatively uncritically to the usage of these terms in the psychological literature; that is, to refer to representations that are characteristically amodal and abstract, possess semantic content, or are capable of being flexibly combined with other representations in the formation of thoughts.

A second debate that I hope I more completely avoid concerns the metaphysical status of consciousness. Everything I say in this dissertation is, I take it, broadly compatible with a range of metaphysical theses about the relationship of the mental to the physical. Hence when I claim, for example, in Chapter 4 that CSTM may be one the constitutive bases of conscious experience, I take this proposal to be compatible with the idea that psychological states are ontologically distinct from physical states.

With these caveats in mind, then, I can now begin my project, and explain what I think is missing from contemporary accounts of consciousness, perception, and memory.

# CHAPTER 1: Short-Term Memory and Theories of Consciousness



## 1.1 – Introduction

One of the key claims of this dissertation will be that most existing models of memory are missing something out, with potential importance for theories of the functional organization of the mind and theories of consciousness. The significance and plausibility of these claims therefore depends upon first giving an overview of the major currents in contemporary memory research. The purposes of this introductory chapter, then, are twofold: first, to give a broad review of current models of memory with a particular focus on short-term memory, and second, to examine how these models have been put to work in debates about the perception-cognition distinction and the theories of consciousness debate.

The first part of the chapter will focus on the former goal, and I begin in 1.2 by describing the history of memory research, focusing on the ideas of Atkinson & Shiffrin (1968), Baddeley & Hitch (1974), and Nelson Cowan (2001). Synthesizing these different approaches, I present a simplified model of memory architecture that I call the 'Sensory-Cognitive Model', which aims to capture what is common to these different frameworks. I go on to consider in more detail in 1.3 and 1.4 two particular components of this model, namely Central Cognition and Sensory Stores. In the latter half of this chapter, I provide a brief introduction to two major debates in the philosophy of mind, the first (in 1.5) concerning the relationship between perception and cognition and the second (in 1.6) concerning the psychological mechanisms of conscious experience, summarizing how these debates relate to the Sensory-Cognitive Model of memory.

## 1.2 – The 'Sensory-Cognitive Model' of memory

Despite the seemingly intuitive distinction between short- and long-term memory, it is only relatively recently that scientists have attempted to give theories that model and account for the difference between the two. A notable and influential early account was given by Donald Hebb (1949), who suggested that short-term memory relied on specific transitory patterns of neural activation, whereas long-term memory relied on more permanent changes involving neuronal growth and regrowth. A more worked out psychological model of this distinction was proposed by Atkinson & Shiffrin (1968). This historically important and still influential picture of human memory posited three distinct kinds of store: high-capacity sensory registers that briefly record incoming sensory information from different modalities; a low-capacity short-term store, allowing for task-specific selection and encoding of information; and a high-capacity long-term store, in which information can be encoded or retrieved over longer intervals (see Fig. 1a).

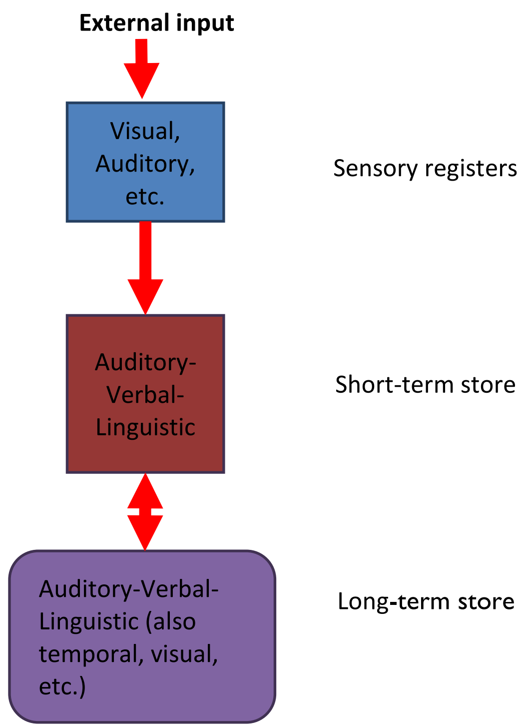



**Fig. 1a**. A schematic illustration of the Atkinson-Shiffrin model (adapted from Atkinson & Shiffrin, 1968)

Though much of this model has been preserved in contemporary theories, the original framework itself struggled to account for a range of subsequent experimental data. For example, the Atkinson-Shiffrin model presupposed that encoding in long-term memory was dependent on initial encoding and rehearsal in short-term memory. If this was the case, one might expect impairments in short-term memory to have massive downstream negative effects on long-term memory. However, one patient (Shallice & Warrington, 1970) displayed drastically impaired short-memory as measured by his ability to recall strings of digits, but was able to learn words and stories quite normally over longer durations. Other work (Craik and Watkins, 1973) showed that the rehearsal of items for longer intervals in short-term memory did not automatically improve the chances that they were later remembered relative to items that had been only briefly rehearsed. Finally, a number of experiments showed that different types of interference tasks affected subjects' short-term memory unevenly; thus, verbal shadowing of a sentence had significantly greater negative effects on subjects' abilities to remember *aurally* presented letter than *visually* presented letters (Kroll et al., 1970).

These results pointed towards a more fractionated and less linear model of memory than Atkinson and Shiffrin had suggested. The Atkinson-Shiffrin model had treated short-term memory as a unitary auditory, visual, and linguistic store, receiving input from sensory stores via a unidirectional transfer of information. Baddeley and Hitch (1974), by contrast, argued that the scientific data could be more readily accommodated by supposing the existence of further types of sensory memory subsequent to initial perceptual processing that could be actively utilized by executive processes in order to maintain data for subsequent tasks.

Intuitively, we often engage in such auditory rehearsal when trying to remember things in the short-term, silently 'playing back' a shopping list or phone number to ourselves to help remember it. Baddeley and Hitch's evidence strongly points towards an active role for this kind of sensory rehearsal in short-term recall. For example, note that *word length* plays an important role in recall, with subjects better able to remember lists of short words than long words (Baddeley, Thomson, & Buchanan, 1975). Interestingly, the advantage for shorter words is abolished if subjects are required to engage in 'articulatory suppression', that is, repeating irrelevant words during the experiment (Baddeley, Lewis & Vallar, 1984). While this would be naturally expected to interfere with the sensory and motor processes involved in rehearsal, it would be much less likely to impair conceptual memory, given that it is hardly an attentionally demanding task.

This sort of evidence, combined with seemingly modality-specific interference effects (for example, the fact that verbal shadowing of a sentence significantly impairs recollection for verbal but not visual material) points to a major role for sensory memory outside of immediate recall. Thus Baddeley and Hitch suggested that two sensory stores, termed the visuospatial sketchpad and the phonological loop, served as distinct buffers for the maintenance of information in what he terms *working memory*, and could be selectively maintained via processes of visuo-spatial or auditory rehearsal respectively(see Fig. 1b).

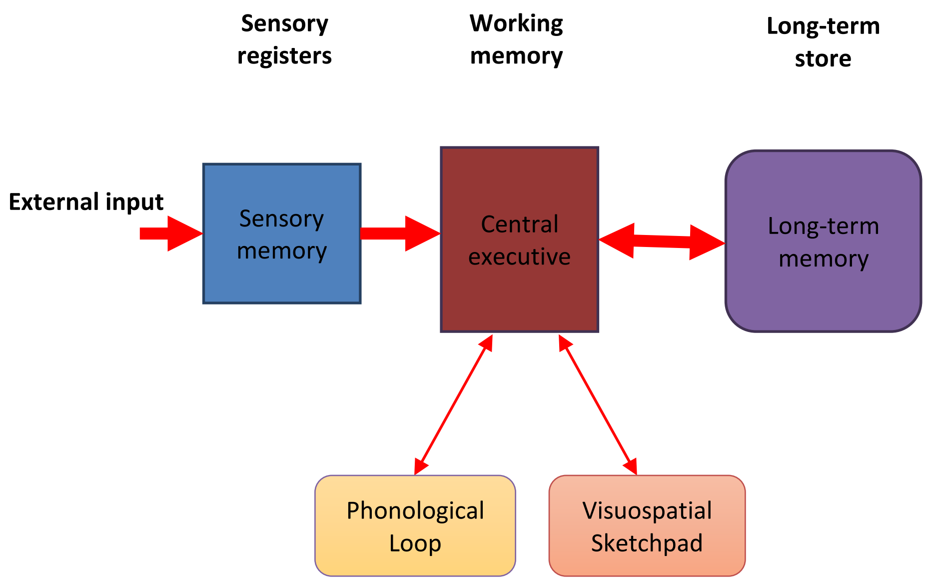



**Fig. 1b.** Baddeley & Hitch's model of memory (adapted from Baddeley & Hitch, 1974)

Baddeley and Hitch's model has in turn been subject to a number of revisions. Their initial account of short-term memory assumed that working memory consisted of nothing other than the visuospatial sketchpad and the phonological loop under the control of an attentional system that did not itself possess any storage capacity. In more recent work, however, they have added the 'episodic buffer', a short-term capacity limited store that can integrate and conceptualize information from sensory forms of memory in single unitary format (see Andrade, 2001:302ff.). A second revision has been prompted by the discovery of other kinds of short-term sensory store that can be used by short-term memory for the maintenance and rehearsal of information in working memory. Specifically, there is now evidence for olfactory, tactile, proprioceptive, and affective forms of stores available to working memory (Dade et al. 2001; Harris et al. 2002; Jeannerod 2006; Mikels et al. 2008). This has led some theorists to suggest that the sensory stores used for rehearsal in working memory may be the very same short-term sensory stores involved in perception. Thus Peter Carruthers claims, for example, that "the 'slave' systems of the central executive are none other than the auditory and visual systems, whose resources can be recruited by executive activity to broadcast and sustain imagistic representations of the relevant sort" (Carruthers, 2014). Other doubts come from Cowan (2001) who suggests that Baddeley's sensory stores are in fact simply activated portions of long-term memory. I will not take a stand on this controversial and largely empirical issue.

Baddeley and Hitch's work emphasized the contribution of sensory forms of memory to the function of central executive processes, something neglected in Atkinson and Shiffrin's original model. A second kind of emendation to the Atkinson-Shiffrin model has concerned the capacity of short-term memory. Whereas Atkinson and Shiffrin (following Miller, 1956) supposed this to be around five to eight items (1968: 112), current evidence points to a significantly smaller capacity of around four items at once (Cowan, 2001). This is not to say that we can only ever recall four words or numbers at a time, of course. There are various ways we can bolster the effective capacity of central executive processing, including 'chunking' (see below) and rehearsing information in sensory stores. However, when chunking is controlled for, and subjects are prevented from rehearsing information in sensory stores (for example, via requiring them to perform verbal shadowing tasks to prevent rehearsal of information), a consistent capacity of 3-4 items is observed.

With the emendations just described in mind, we are now in a position to lay out a broad functional architecture that captures the main insights of the current leading theories. One of my goals in doing this will be to establish a simple set of terminology to refer to different stages of processing. This is essential because there is an unfortunate tendency to use terms like 'working memory' to refer to different mechanisms. Thus Baddeley uses the expression to refer to the whole assembly of sensory and amodal mechanisms involved in executive processing together with their attentional controller. However, this is perhaps not ideal, given that it remains controversial whether these sensory mechanisms are truly distinct from the sensory buffers utilized in short-term perceptual memory, and the relation between attention and other mental faculties is controversial.[1] Moreover, in contemporary philosophical and scientific literature, the term 'working memory' is more commonly used by contemporary theorists to refer just to the single multi-modal capacity-limited informational store that Baddeley terms 'the episodic buffer' and Cowan calls 'the focus of attention'. Reflecting contemporary usage, then, I will use the term 'working memory' in this this latter sense rather than Baddeley's original usage.[2]

With this in mind, I am now in a position to describe the Sensory-Cognitive Model. This incorporates three things. The first is a set of short-term, time-limited sensory stores (such as iconic memory), utilized in perception and perhaps in the rehearsal of perceptual information. The second is an executive faculty used for high-level tasks including the maintenance and reporting of multimodal and conceptual information. Following Carruthers (2014), I term this faculty Central Cognition. While this faculty may make use of sensory

[1] See Appendix 1 for a brief review of the data concerning attention and central cognitive processes.
[2] For the purposes of this dissertation, I use the term "working memory" to refer just to states that are reportable and available for voluntary action (see the discussion of cognitive access, below). There is, however, some controversial new research that points to the existence of a further kind of working memory that is seemingly *unconscious*. The existence of this unconscious working memory is still hotly debated, however, so I will not discuss unconscious working memory in the main body of this dissertation, but a summary of current data is presented in Appendix 3.

stores, in the final case, its performance is limited by a single central storage mechanism that I will term working memory.[3] The final component of the Sensory-Cognitive model is a long-term store with a large capacity and extended duration (see Fig. 1c).

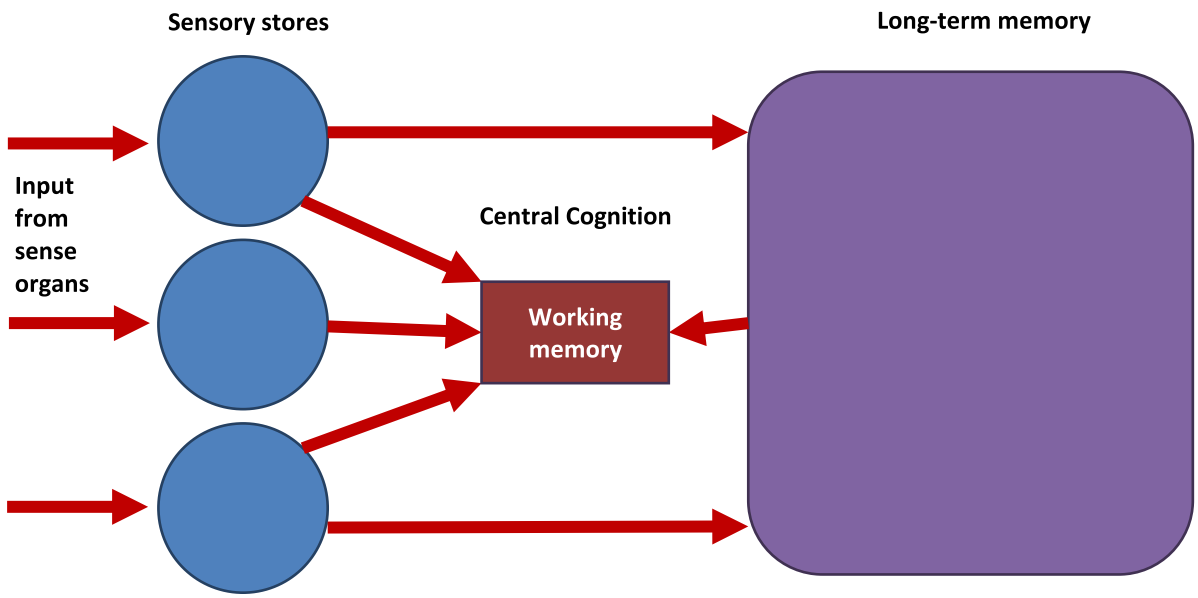



**Fig. 1c.** A highly schematic depiction of the 'Sensory-Cognitive Model' illustrating the forms of short-term memory. Note that given the ongoing controversy surrounding the phonological loop and visuospatial sketchpad, I have not included them in this diagram.

As should be evident, this is in many respects similar to the original model of Atkinson & Shiffrin. I will refer to this broad functional architecture as the 'Sensory-Cognitive Model'. Note that it has been deliberately cast in fairly vague terms to accommodate a variety of distinct approaches, and in particular, to accommodate the models of both Baddeley & Hitch (1974) and Cowan (2001). It is nonetheless specific enough to allow us to spell out ordinary human memory processes in a little more detail.

[3] As noted above, this should be identified specifically with Cowan's notion of the 'focus of attention' and Baddeley's 'episodic buffer'.

## 1.3 – Central Cognition

Having given a broad sketch of the functional organization of memory, I now wish to say a little more about the two short-term mechanisms in this model, namely Central Cognition and sensory stores. Consider Central Cognition first. This is a robust, flexible, but highly capacity-limited executive processing mechanism incorporating a dedicated buffer for storing information which I have termed working memory. It is recruited for a wide range of tasks involving learning and voluntary action, but is usually experimentally studied via complex attentionally-demanding tasks, such as memorizing lists of unfamiliar numbers or words (Broadbent, 1975) or detecting changes in an array (Vogel, Woodman, & Luck, 2001). Although it may draw upon the resources of various modality-specific sensory buffers as discussed below, Central Cognition allows for the combining of information in working memory from different modalities and their conceptualization in non-modality specific terms (Baddeley & Hitch 2010), though it remains controversial whether it ever incorporates wholly amodal information (Carruthers, 2015a).

Note that while it is common to mark distinctions between different forms of working memory such as visual working memory and olfactory working memory, there is no particularly strong empirical reason to think that these constitute fundamentally distinct psychological mechanisms (see Block 2007 for some discussion). Rather, I will proceed on the hypothesis (recently supported by the data of Salmela, Moisala, & Alho, 2014) that Central Cognition ultimately depends on a single multimodal buffer that in turn draws information from different lower-level modality-specific sensory stores.

Critical early work on Central Cognition and its capacity limits was presented in the classic paper "The Magical Number Seven, Plus or Minus Two", in which George Miller claimed that both short-term recall and perceptual judgments were constrained by a processing limit of around seven items at once. Subsequent research suggested that matters were considerably more complicated, however, and different types of items (words and digits, for example) are remembered more or less easily. Another confounding factor comes from the fact noted earlier that individuals can expand the effective memory capacity of Central Cognition by rehearsing information in short-term sensory stores. However, perhaps the most important confounding factor in assessing short-term memory capacity concerns what is known as chunking. A common example of this might be remembering a phone number by breaking it into identifiable strings. For example, the number 555-313233 can be easily remembered as consisting of three fives followed by counting up from thirty-one. In one famous case of chunking, an individual was able to learn to remember eighty digits via breaking up the numbers into strings of five- to ten digits at a time (Ericsson, Delaney, Weaver, & Mahadevan, 2004).

Nonetheless, when experiments are designed to control for chunking (as well as the employment of sensory forms of memory), there is a surprisingly high degree of regularity in people's capacity to store distinct pieces of information at a time, namely a limit of three-to-five items at once. [4] This holds true regardless of whether the relevant information to be retained is a high-level semantic representation (as in a digit or word) or simpler sensory information (as in a color patch or tone). A summary of some of the evidence for this capacity limit is included below (Fig. 1d).

[4] Note that there is evidence that interpreting this effective limit on Working Memory capacity in terms of a finite number of 'slots' may be at least somewhat inaccurate. Cutting-edge research on the capacity of working memory suggests that the observed 4-item limit may reflect the challenge of recovering information from 'noisy' signals. See Appendix 2 for further discussion.

| Study | Stimulus type | Procedure | Memory estimate |
|---|---|---|---|
| Baddeley et al. 1974 | Verbal (Monosyllabic words) | Auditory presentation with articulatory suppression | 3.5 |
| Darwin et al. 1972 | Auditory (words) | Partial report | 4 (after 4s delay) |
| Pylyshyn et al. 1994 | Visual objects | Multiple object tracking | 3-5 |
| Sperling 1960 | Visually presented letters | Whole report | 4 |
| Broadbent 1975 | Word lists | Errorless recall | 3-4 |

**Fig. 1d.** Some evidence for a 4-item capacity limit in short-term memory tasks (see Cowan, 2001).

The notion of Central Cognition and its limited working memory store has, under its different guises, proven to be an extremely fertile theoretical posit. For example, measures of individuals' working memory capacity seem to predict a wide range of other cognitive capacities, including reading ability, problem solving, and general intelligence (Conway, Kane, & Engle, 2003; Spearman, 2005). Additionally, evidence has been presented that working memory training enhances performance on a wide range of other executive tasks, including attentionally demanding perceptual abilities (Jaeggi et al., 2008; Schwarb, Nail, Schumacher, 2015), though this line of research remains controversial. There has also been investigation of working memory in non-human animals, examining, for example, the ability of rats to combine visual and olfactory cues in problem solving (Bratch et al., 2016).

While it is widely accepted that some faculty along the lines of Central Cognition exists, it is also subject to a number of outstanding controversies. Some of these controversies (such as the question of whether *encoding* and *maintenance* in working memory are subserved by the same attentional controllers) are not directly relevant to the current project. Three issues in particular, however, are more relevant to the concerns of this dissertation, namely (i) the relation between attention and working memory, (ii) the question of whether working memory uses 'slots', and (iii) the existence of unconscious forms of working memory. Given the somewhat complex empirical issues surrounding these topics, I will not discuss them in the main body of this chapter, but instead address them individually in Appendices 1-3.

## 1.4 – Sensory Stores

I now turn to discussion of the notion of sensory stores. This category encompasses several distinct forms of memory that are characterized by their relatively high capacity, brief duration, and susceptibility to disruption by the presentation of new information in the relevant sensory modality.

Evidence for the existence of sensory stores, as distinct from both Central Cognition and long-term memory, comes from two major kinds of data. The first sort of data comes from experiments in which subjects undergo 'sensory overload'. As noted, when dealing with distinct items that that cannot be easily chunked, subjects struggle to recall more than three or four items from a given stimulus, even when tested immediately after presentation. However, in certain paradigms, subjects presented with complex arrays of different items can be cued to reliably report on some specific *subset* of items in the array, even when this cue occurs *after* the stimulus has been removed. This in turn suggests that subjects can retain a complex representation of the stimulus even after it is removed, and that their reports reflect only a small part of the total information that was briefly available.

The most famous example of a sensory overload experiment is George Sperling's work on iconic memory and partial report paradigms. In the original Sperling experiment, subjects were briefly shown a 3x4 matrix of alphanumeric characters (Sperling, 1960; see Fig. 1e). After the stimulus was removed, if subjects were immediately cued to report on any given row, they were able to successfully report almost all the contents of that one row (3-4 items), though not other rows. This suggests that subjects have brief access (for up to a second) to a fragile memory store that encodes 9-12 items from the initial stimulus.

**The Sperling Test**

Reported characters are indicated with a glow, cued row indicated by an arrow.

**Whole report:** subjects are presented with a matrix of alphanumeric characters for 15-500ms. When asked to report all of the characters they can remember, subjects can report on average 4-5 /12 characters, an accuracy of 33-40%.


**Partial report:** However, if subjects are cued to report just on *one particular row* (top, middle, or bottom) subjects can correctly remember 3-4 characters from that row, an accuracy of 75-100%, but few if any others. This effect persisted for up to 1000ms in Sperling's original experiment.


**Fig. 1e.** The Sperling Test (1960).

Sperling's work showed that subjects were able to selectively report on items that were cued, even where the cue followed stimulus removal, but were unable to report on non-cued items. This strongly suggests a distinction between the robust form of short-term memory that is required for report (in present terms, working memory) and a more fragile visual buffer.

A second source of evidence for sensory memory comes from the different ways in which multiple tasks can interfere with subjects' ability to retain information in the short-term. As noted above, working memory seems to have a limited capacity of around 3-4 items, and when chunking is properly controlled for, subjects struggle to keep more than this number of items clearly in mind. However, their performance can be significantly improved if the items they are required to recall are spread across distinct sensory modalities (see, e.g., Baddeley 1986; Frick 1984; Greene 1989; Shah & Miyake 1996).

While Sperling's iconic memory is the most famous example of a sensory store, there are a number of other well established forms of sensory memory. For example, Ilja Sligte has demonstrated a form of visual sensory memory termed Fragile Visual Short-Term Memory (fVSTM) which persists significantly longer than iconic memory, and seems to have an even higher capacity, with an upper limit of at least 32 objects (Sligte, Scholte, & Lamme, 2008). There is also at least one non-visual form of sensory memory, specifically in audition. This form of memory, dubbed 'echoic memory', is similar to iconic memory in having a very large capacity, but persists for slightly longer durations (Darwin, Turvey, & Crowder, 1972).

There are two features of sensory memory that are worth briefly dwelling on. The first is that, particularly in the case of visual forms of sensory memory, it is readily disrupted by further stimulation in the relevant sensory modality, such as a pattern mask in the case of iconic memory (Averbach & Coriell, 1961; note also that Saults & Cowan, 2007, show overwriting effects in echoic memory). The same is true of fVSTM, though it is a little more resilient, requiring location-specific masking to 'overwrite' previously presented information (Sligte, Scholte, & Lamme 2008; Pinto et al. 2013). A second important feature is that sensory forms of memory probably do not encode objects in respect of higher-level semantic properties such as conceptual identity, but instead just represent sensory properties such as color and shape. Considering the case of iconic memory for example, subjects in partial report paradigms do not exhibit superiority when cued to report just on items belonging to a given semantic category, such as ‘numbers’ or ‘letters’ (Sperling, 1960; von Wright, 1970). The data on the nature of encoding in Fragile Visual Short Term Memory is more complex; for example, it allows in some cases for binding of different visual features to a single object (Landman, R., Spekreijse, H., & Lamme, 2003), and Lamme (2006) suggests that it may allow for representation of moderately complex content such as faces, for example. However, there is no evidence I am aware of that fVSTM can encode multimodal or high-level conceptual or semantic information. This is not surprising given that fVSTM is hypothesized to rely on recurrent processing based in low- and mid-level visual areas (Lamme, 2010).

## 1.5 – Applying the Sensory-Cognitive Model I: Perception and Cognition

I have now given an overview of the major stages of processing invoked by current theoretical work on memory. In the next two parts of this chapter, I wish to briefly mention how these empirical models of memory may relate to two philosophical debates, beginning with questions about the perception-cognition distinction and moving on to the theories of consciousness debate.

The question of how to distinguish perception and cognition is a long-standing area of dispute in philosophy of mind, but has been a topic of particular interest in the last few decades, featuring prominently, for example, in the work of Dretske (1981), Fodor (1983), Burge (2010), and Block (2014a, 2016), though some theorists (such as Clark, 2013) still doubt whether the distinction is a robust one.

Note that the question is not simply a matter of how to distinguish *perceptual experience* from *conscious thought*. Though the two questions are intimately connected, one might nonetheless hold, for example, that perception and cognition are importantly distinct psychological kinds while also maintaining that what we call ‘perceptual experience’ and ‘conscious thought’ are both subserved by a single strictly cognitive mechanism.[5] The most basic question is instead whether there is some fundamental psychological distinction to be drawn between two kinds of processing in the brain that might reasonably capture key aspects of the pretheoretical distinction between perception and cognition. Independently of questions about consciousness, this is important for understanding the architecture of the mind (as discussed in Fodor, 1983) and issues concerning the extent to which mental states like beliefs and desires might influence sensory processing (as discussed by Pylyshyn, 1999).

The debate concerning the perception-cognition distinction connects with issues in short-term memory research in several ways. First, several theorists have proposed that perception and cognition, like sensory memory and Central Cognition, are distinguished by their *formats*. Thus Dretske claims that "…our perceptual experience... is coded in analog form and made available to something like a digital converter… for cognitive utilization" (1981: 153). Proposals along broadly similar lines have been developed in the work of Tyler Burge (2010, 2014b). Burge follows Dretske in taking perception and cognition to differ fundamentally in format, with perception but not cognition possessing an iconic format (2010b).

The key point to note here is that considerations of format *also* apply to mechanisms of short-term memory. Sensory forms of memory, as noted earlier, encode objects just in respect of low-level modality-specific properties like shape and color, whereas representations in working memory proper are represented in a "common multi-dimensional code" (Baddeley, 2000). Indeed, the connection to work on short-term memory is explicitly invoked by many of the authors mentioned, with both Fodor (2007) and Dretske (1981) pointing to Sperling's work on

[5] For example, this may be one way of interpreting David Rosenthal's Higher-Order Thought theory in light of his account of mental qualities (Rosenthal 2005, 2010).

iconic memory as an example of strictly perceptual analogue processing.

A second way in which these issues connect to work on memory concerns the *capacity* of different stages of processing. Thus it is often held that perception is "rich", processing large numbers of objects and properties in parallel, whereas cognition can process relatively few things at once. Thus Dretske, for example, says as follows.

> Our sensory experience embodies information about a variety of details that, if carried over *in toto* to the cognitive centers, would require gigantically large storage and retrieval capabilities. There is more information in the sensory store than can be extracted, a limit on how much of this information can be exploited by the cognitive mechanisms.

A similar point is made by Fodor (2007), who takes the lack of capacity limits for certain tasks ("item effects") to be evidence of iconic perception.

A partially capacity-based distinction between perceptual and cognitive processes has also been more recently defended by Ned Block (2007), who claims that "perceptual consciousness overflows cognitive access" (2011a). This difference in capacity provides a further way of carving up the perception-cognition boundary, albeit one that is closely linked to considerations about format: for Dretske and Fodor (and perhaps for Block), part of the *reason* perception has a large capacity is that it can exploit analogue or iconic forms of information in which the coding of individual features is less computationally demanding than is the case for digital or conceptual formats.

Again, many of the defenders of this distinction (including Fodor, Dretske, and Block) explicitly appeal to memory mechanisms in the form of working memory and sensory memory in arguing for the relatively high capacity of perception as opposed to cognition, and more broadly in distinguishing cognition from perception. Note, for example, Dretske's reference to the "sensory store" in the above quotation: this is not merely an idle borrowing of terminology from psychology, but he explicitly has in mind mechanisms such as iconic memory, and refers to the work of George Miller discussed earlier as evidence for the limited capacity of cognition. Similarly, Fodor and Block, in claiming that the capacity of cognitive mechanisms is limited relative to the capacity of perception, appeal to Sperling's work on high-capacity iconic memory.

While it would be oversimplifying matters to say that any of these authors take these memory mechanisms to be *constitutive* of perception or cognition, reference to short-term memory models provides us with a further empirical tool for assessing the differences between these two parts of the mind. If the Sensory-Cognitive model is correct, the perception-cognition distinction may be able to be neatly mapped onto distinctions between the two main forms of short-term memory. On the other hand, to the extent that the Sensory-Cognitive Model is incomplete (as I will argue in the next chapter), these accounts of the perception-cognition distinction may in turn require revision to accommodate challenging cases, particularly categorical perception. I will return to these concerns in Chapter 3.

## 1.6 – Applying the Sensory-Cognitive Model II: Theories of Consciousness

A similar and closely related debate in which mechanisms of memory may prove to be of relevance is the theories of consciousness debate. The central question in the theories of consciousness debate, put simply, is why some mental states are conscious and others are not. A theory of consciousness thus aims to provide a set of necessary and sufficient conditions for a mental state's being conscious. These conditions may spell out the psychological properties involved in a state's being conscious (such as global availability, or being the target of a higher-order thought), or their neural properties (such as exhibiting a 40Hz firing rate), or more typically, a combination of the two (as in Prinz, 2012). Note that the theories of consciousness debate is thus somewhat distinct from the "hard problem of consciousness" (Chalmers, 1995), since it need not attempt to answer *why* consciousness exists in the first place, instead spelling out conditions that describe *when* it arises in a given system.

Knowing the conditions under which consciousness emerges would be of vast practical and scientific benefit. Of course, most of the time, we can reliably ascertain when someone is undergoing a conscious state simply by asking them. But matters are not always so simple. Patients in persistent vegetative states are non-responsive to verbal questions, but it is possible that they nonetheless undergo conscious experiences; at present, we have few ways to answer this question. We also have little grip on which non-humans animals undergo conscious states, and how broadly consciousness is distributed in nature.[6] These latter concerns also have ethical import: on the assumption that consciousness is a prerequisite for suffering, our dealings with animals could be made significantly more humane if we knew which animals could undergo experiences of pain and suffering (see 5.6 for further discussion). A theory of consciousness could similarly be applied to assess consciousness in neonates and fetuses, and even perhaps to questions about intelligence in artificial systems.

One critical distinction for the theories of consciousness was proposed by Ned Block (1995) in order to address what he took to a failure of leading scientific models of consciousness to properly engage the underlying phenomenon. Specifically, Block proposed a distinction between *phenomenal* consciousness (or p-consciousness) and cognitive access (which Block refers to as access consciousness in his earlier work, with broadly the same meaning).

Consider cognitive access first. Block suggests that being cognitively accessed amounts

[6] Note that there is a distinction to be drawn here between theories of state consciousness (the conditions under which a given state is conscious) and creature consciousness (the conditions under which a given individual is conscious). I will not explore this distinction in this dissertation, though I will note that I broadly endorse the 'phenomenal field' account of Bayne (2007) according to which creature consciousness can be spelled out in terms of two components, namely the presence of a conscious state within a given unified subjective perspective or 'field'.

to a mental state's being "poised for use as a premise in reasoning…. poised for rational control of action, and… poised for rational control of speech." Thus subliminally presented images or unattended and unreportable stimuli are not cognitively accessed, while any item that a subject can immediately report can be assumed to have been cognitively accessed (1995: 231).

Whereas Block's notion of cognitive access is stipulative, his concept of phenomenal consciousness is descriptive, and aims to capture something that is pretheoretically obvious, namely the concept of what it's like to have an experience. Indeed, Block states that he "cannot describe P-consciousness in any non-circular way". However, he suggests that the term "experiential properties" as a synonym, and points to a familiar range of examples in the form of the subjective character of sensations, feelings, desires, and so on. A mental state is phenomenally conscious if there is something it is like for a subject to undergo that state – that is, it has some subjective character. Thus when I consciously see red, or smell coffee, or feel a flash of anger, in each case there is some way that the experience feels to me.

Block himself has been at pains to show that we should not simply *assume* that cognitive access and phenomenal consciousness are one and the same thing. While an account of the mechanisms of cognitive access is an important goal for psychology, what we ultimately want from a theory of consciousness is an account of what it is in virtue of which a given mental state is phenomenally consciousness.

The theories of consciousness debate closely relates to issues in memory research because cognitive access is taken by most theorists to require encoding in working memory.[7] As Block notes, for example, all "first-person reports [about] phenomenal consciousness [are]

[7] For an important exception, see Carruthers, 2015b. While Carruthers grants that "items [must] be retained in working memory for subsequent report", he suggests that the notion of cognitive access should be taken to include not just items that are *immediately* available for report, but all items that are globally broadcast and *available* to working memory. However, it is open for debate whether this amounts to a kind of subtle redefinition of Block's notion of cognitive access in somewhat broader terms than originally intended. See 4.7 for more discussion.

filtered through our cognitive access to it via perceptual working memory" (Block, 2014b). Similarly, Prinz (2007), responding to Block, summarizes cognitive access as ranging just over those "perceptual states that have been accessed by working memory". As far as most participants in the debate will be concerned, then, we can recast the question of whether phenomenal consciousness constitutively requires cognitive access by asking whether phenomenal consciousness requires encoding in working memory.

Spelling out the debate in these terms, then, we can offer a tripartite scheme for characterizing the main ways in which contemporary theorists of consciousness have responded to Block's challenge (cf. Prinz, 2007). First, many theorists (notably Dehaene, 2014, Kouider et al. 2010, and Baars, 1988) have claimed that some form of actual cognitive access is, as a matter of empirical fact, necessary for consciousness.[8] This is also the position adopted by both Baddeley (2003: 836) and Cowan (2001: 91). Cowan, for example, claims that only "information in the focus of attention is available to conscious awareness and report", while Baddeley, describing the role of the episodic buffer, states that it is "regarded as a crucial feature of the capacity of working memory to act as a global workspace that is accessed by conscious awareness along the lines suggested by Baars." If such views are correct, we can rule out, among other things, the possibility that a subject could be in a conscious state that was not immediately available for report. I will term those who endorse this position *Access theorists*.

Second, there are those theorists, among them both Ned Block and Victor Lamme, who have embraced the dissociation between phenomenal consciousness and cognitive access and

[8] These theorists all subscribe to versions of Global Workspace Theory (discussed in Chapter 4) and thus primarily make reference to global broadcast rather than encoding in working memory. However, as Baddeley's comments quoted below suggest, the two notions are closely connected; Dehaene, for example, states that one of the functions of the broadcasting in the global workspace is to "provide... a working memory space" (2011: 63). In his most recent work, Dehaene has identified the contents of consciousness with the activated portion of working memory specifically (Dehaene, 2016).

accessibility at an empirical as well as conceptual level, and have attempted to ground theories of consciousness in non-cognitive mechanisms such as recurrent feedback loops in early visual areas. I will refer to those who take this approach as *Access-independent theorists*.

A final group of theorists, including Peter Carruthers (2011), Michael Tye (1995), and Jesse Prinz (2012), offer theories that are compatible wiith Block's claim that phenomenal consciousness does not depend on actual cognitive access while nonetheless maintaining that all conscious states must at least be *cognitively accessible* to a subject.[9] Thus Prinz suggests that "consciousness arises when intermediate-level perception representations *are made available* to working memory via attention" (2005; emphasis added). They thus allow for a functional distinction between conscious experience and cognitive access, while preserving a role for broader cognitive properties (namely, cognitive accessibility) in determining whether or not a given state is conscious to begin with. I will refer to theorists who endorse this view as *Accessibility theorists*.

This tripartite distinction between Access, Accessibility, and Access-independent theories is somewhat crude, and arguably provides a poor fit for some important approaches to consciousness such as Tononi's Information Integration Theory and Rosenthal's Higher-Order Thought account (Tononi & Koch, 2015; Rosenthal, 2005. See Appendix 4). Nonetheless, it provides a useful way to frame many of the leading theories of consciousness in terms that are tractable within the framework of current research into short-term memory.

However as intimated earlier, in the next chapter, I will explore evidence suggesting that we should amend the Sensory-Cognitive account of short-term memory to include a further form

[9] In more recent work, however, Carruthers has characterized himself as a cognitive access theorist; see previous footnote.

of processing. This will put pressure on the tripartite distinction just described, and may in turn create theoretical space for a new account in the theories of consciousness debate that I term the Workspace-Plus account. I will return to this suggestion in Chapter 4, where I spell out the theories of consciousness debate in more detail and outline this new approach.

## 1.7 – Conclusion

My goals in this chapter have been twofold. The first was to build a shared framework that captures the leading insights from different current approaches to memory, which I have termed the Sensory-Cognitive Model. This core, I suggested, consists of two main kinds of short-term stores, namely sensory stores and working memory. My second goal was to illustrate two disputes in which the components of this model have been influential, namely questions about the relationship between perception and cognition and the theories of consciousness debate.

In the next chapter, however, I will put the Sensory-Cognitive Model under the spotlight, examining a challenge that it faces from contemporary research. In particular, I will consider the proposal that the Sensory-Cognitive Model is incomplete, and I will examine evidence for a further stage of short-term memory in the form of Conceptual Short-Term Memory. In turn, I suggest, this opens up new possibilities in both of the debates just mentioned, which I explore further in Chapter 3 and Chapter 4.

# CHAPTER 2: Conceptual Short-Term Memory



## 2.1 – Introduction

The Sensory-Cognitive Model described in the previous chapter captures the core claims of most leading theories of memory, and can adequately accommodate much of the last six decades of memory research. Nonetheless, it should be borne in mind that memory research as a whole is still a relatively young field, and a number of important questions remain, especially in regards to how memory is related to attention and how information is actually encoded in working memory (See Appendix 1 and 2).

However, I now wish to examine evidence for the claim that a more fundamental revision should be made to the Sensory-Cognitive Model as provided in Chapter 1. Specifically, I will describe some experiments recently marshalled by psychologists in support of the claim that there is an important intermediate stage in processing in the form of *Conceptual Short-Term Memory,* a short-term cognitive buffer in which high-level semantic information about presented sensory stimuli (such as their categorical identity) is briefly retained.

In proceeding, I should note that in much of this chapter I will be presenting and evaluating evidence for specific theses in empirical psychology. My arguments here are of course not intended to settle the live psychological debate about the existence of CSTM: its status remains controversial among psychologists, and ultimately these questions will be settled in the laboratory rather than by philosophers. I nonetheless take there to be significant philosophical value in assembling the evidence for CSTM and spelling out its putative characteristics in order to examine its potential consequences for questions in the philosophy of mind. After all, as noted in the previous chapter, philosophers have frequently appealed to short-term memory mechanisms in support of various theses about perception, cognition, and consciousness, hence the possibility that there is an important form of memory being left out could be highly relevant for these debates.

The chapter proceeds as follows. I begin in 2.2 by giving a very general specification of CSTM and how it differs from both sensory stores and working memory. In sections 2.3-2.5 I present a range of empirical evidence that has been taken to support the existence of CSTM. In 2.6, I spell out in more detail the CSTM hypothesis, and show it may point to a possible revision to the Sensory-Cognitive Model accordingly. Finally, in 2.7, I briefly examine the relation between CSTM and the experimental data on unconscious semantic processing. Note that in this chapter, I will attempt to remain neutral on the question of whether CSTM is conscious. However, these considerations will be examined in more detail in later chapters.

## 2.2 – Enter CSTM

As discussed in the previous chapter, the Sensory-Cognitive Model takes there to be two main kinds of short-term memory store, namely sensory stores and working memory. Broadly speaking, sensory forms of memory are fragile, have a high capacity and brief duration, and seem to encode objects in visual or other dedicated sensory formats in respect of relatively low-level such as color, shape, and size. Working memory, by contrast is more robust, has a limited capacity, and can encode its contents in a unitary format allowing for representation of fairly high-level properties.

I now wish to present evidence for the claim that there is a third store, conceptual short-term memory (CSTM), that is distinct from these forms of memory. CSTM was initially posited by Mary Potter, and she characterizes it as "a mental buffer in which current stimuli and their associated concepts from long term memory… are represented briefly, allowing meaningful patterns or structures to be identified" (Potter, 2012).[10] CSTM can be distinguished from sensory forms of memory in two main respects. The first is that it seems capable of encoding objects in respect of a wide range of high-level properties such as 'picnic', 'wedding', or 'motorbike helmet'. The second is that, unlike iconic memory or fVSTM, it largely seems *not* to be vulnerable to overwriting by the presentation of new information. It can also be distinguished in turn from Central Cognition via its relatively larger capacity and its seemingly brief duration.

The addition of CSTM, I will argue, may constitute a critical and fundamental emendation to the Sensory-Cognitive Model, and as such is of considerable interest to philosophers and cognitive scientists. The specific hypothesis I wish to explore is that CSTM may serve as an intermediary between perception and cognition that functions to enrich sensory inputs by conceptualizing them in respect of higher-level categories prior to access by Central Cognition (see Chapter 3), thereby allowing subjects, for example, to pre-select representations for attentionally-demanding central processing based on its relevance to their goals and interests.

[10] Note that Potter sometimes (e.g., 2012) talks of CSTM as a *type* of working memory, albeit one with quite distinctive features. In light of the Sensory-Cognitive model described in Chapter 1, and for the sake of clarity, I will continue to use the term 'working memory' just to refer to the more strictly capacity-limited store of Central Cognition.

I take this suggestion further in Chapter 4, arguing that this process of conceptualization may constitute the lower bound of conscious experience, being responsible, in Potter's terms, for "the unreflective understanding that is characteristic of everyday experience" (Potter, 2010).

I will return to these more theoretical proposals in the later chapters. For now, my goal is to present and evaluate the main psychological evidence that has been advanced in support of CSTM, and to show how this distinguishes it from the two short-term memory stores of the Sensory-Cognitive Model. This evidence comes from three main sources, as I will now describe. The first source of evidence (given in 2.2) is a series of experiments in which subjects seemingly retain very large amounts of semantic information from rapidly presented words or pictures. The second source of evidence (given in 2.3) consists of data from rapid parallel visual search tasks, in which subjects seem to be sensitive to the semantic properties of large numbers of stimuli. The third source of evidence (in 2.4) is the attentional blink phenomenon, and the finding that high-level semantic processing of supraliminal stimuli can occur even in the absence of access by Central Cognition. Together, I suggest, this evidence amounts to a strong case for CSTM.

## 2.3 – Evidence for CSTM I: RSVP

As noted earlier, much of the evidence for CSTM comes from the work of Mary Potter, whose research into short-term memory has spanned several decades. Much of Potter's work involves a technique known as Rapid Serial Visual Presentation (RSVP), a process in which subjects see a series of words or picture presented one after another in rapid sequence.

Some initial evidence for a semantic buffer distinct from Central Cognition came from work in which subjects seem to have brief but fleeting awareness of large numbers of rapidly sequentially presented words (Potter, Kroll, & Harris, 1980). It was discovered that subjects briefly presented with the words from a twelve-word sentence one at a time can accurately remember every word they see, but cannot do so for a similarly presented list of twelve unrelated words. This suggests that, at some level, subjects are able to access the meanings of those words very quickly in order to establish whether there are appropriate semantic and syntactical connections between them. If the relevant connections are present, the sentence can be retained in Central Cognition via chunking mechanisms, but in cases where those connections do not exist, the information cannot all be stored in working memory and consequently quickly degrades.

Another experiment of Potter's provided further evidence for the idea that this conceptual information is only very briefly available, and rapidly lost unless consolidated in working memory (Potter, Stiefbold, & Moryadas, 1998; see Fig. 2a). Subjects saw a sentence rapidly presented one word at a time, each word being displayed for 133ms. At one point in the sentence, subjects briefly saw a pair of words shown simultaneously, only one of which was contextually appropriate. Their task was to pick out the contextually appropriate word and repeat the whole sentence, a task they performed well at. However, they were frequently unable to recall the word whose meaning was rejected, even though its semantic content must have been somehow accessed in order for subjects to prefer the other word as more contextually appropriate.[11]

[11] As my references to 'at some level' should suggest, I recognize that these experiments do not necessarily show that subjects were consciously aware of *all* of the stimuli presented. It may be the case that, for example, selection of the contextually more appropriate word in Potter et al. 1998 was performed by subpersonal or otherwise unconscious mechanisms. Nothing of great importance will hinge on this consideration in this chapter. However, I find it a phenomenologically compelling hypothesis to suppose that subjects may indeed have been briefly aware of all presented words, even those they forgot. Consider, for example, the experience of driving along the highway while glancing at passing billboards. In such circumstances, it certainly seems as though we may become fleetingly aware of the meaning of the words we see, even though they are almost immediately forgotten.

```
                            pencil
Maggie carried the kitten in a ****** to her house
                            basket
```

**Fig. 2a** (from Potter et al. 1998). Subjects saw a series of words like those below, and had to choose the semantically more appropriate of two words before repeating back the whole sentence.

These experiments seem to demonstrate that high-level semantic information about visual stimuli is accessed extremely rapidly and almost immediately forgotten. However, it remains possible that this is explicable in terms of existing forms of memory, such as rapid serial encoding in Central Cognition. More recent work of Potter's is harder to explain away in these terms, however (Potter et al., 2014). In one crucial experiment, subjects were shown 6-12 sequential images (which they had not previously seen) for durations of 13, 27, 53, or 80ms (see Fig. 2b). They were given a target description (for example, 'wedding' or 'flowers') 900ms before or 200ms *after* presentation of the images, and asked to say whether any of the images they saw matched the description.

The results were striking: despite the brief durations, large numbers of presented items, and cueing after stimulus removals, subjects were above chance in all trials. However, their degree of performance was influenced by a few factors. For example, the longer the initial duration of the stimuli, the more likely subjects were to correctly detect a target. Additionally, in trials in which they were given the description in advance, subjects were able to detect presented stimuli at a higher rate of accuracy. Subjects who were shown just six images rather than twelve performed better in all measures, but again, subjects performed at well above chance levels on all measures. Despite their excellent performance on these objective measures, the subjects only reported a rapidly changing, very short sequence of colors and shapes (save for the last picture which subjects were more confident of having seen, but which they were not tested on).

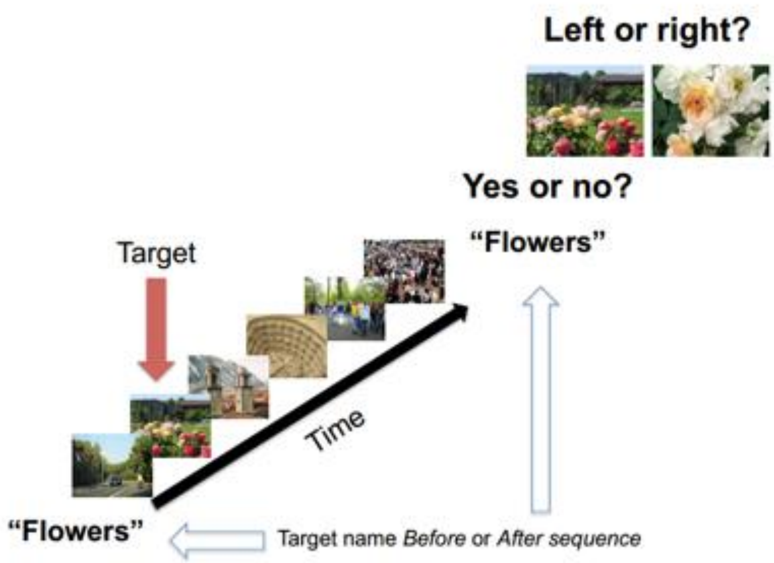



**Fig. 2b (from Potter et al. 2014).** Sequences of 6-12 images were rapidly presented. Subjects were given a target description 900ms before or 200ms after presentation and instructed to say whether they saw a picture matching that description. In trials where a picture was presented, they performed a subsequent recognition task.

Some of the trials also incorporated a second task. Having been asked to identify whether they had seen an image matching a given description, subjects were subsequently shown two images matching that description and given a forced choice task in which they had to indicate which of the two images was the one actually presented. They were given this task even on trials where they had not successfully detected the image. Subjects' performance in this second task was closely linked to whether they had made a successful detection of the target under a given description. For subjects who failed to detect the initial stimulus, "their forced choice was near chance, suggesting that the visual features of unidentified pictures were not retained" (Potter et al. 2014: 276).

We can begin to draw a rough picture about what must be happening in this experiment. First, in order to be able to reliably perform detection and recognition tasks subsequent to presentation of the stimuli, subjects presumably had to retain information about all or most of the

6-12 presented images. Moreover, in order to be able to *use* this information to say whether any of the images matched a given target description, subjects must either have already stored semantic information about each image or otherwise be able to rapidly derive it.

So much should be relatively clear. What is much more open for debate is *which* memory mechanisms underpin subjects' performance. Potter's own assessment is that this result is best explained in terms of a dedicated memory mechanism in the form of CSTM. Building on Potter's suggestion, I think one attractive interpretation is that subjects briefly stored information about all stimuli, encoding both high-level semantic information (so as to be able to detect whether any stimulus fitted a given description) bound to some lower level information (so as to enable accurate recognition). This was momentarily available, such that, when cued with a target description, subjects were able to tap the information in CSTM via Central Cognition to give accurate reports, as well as being able to recognize the relevant image if it was presented to them. In cases where there was no encoding in CSTM, subjects were both unable to detect the image according to its description, and were at chance in recognizing it.

This seems to be the most straightforward explanation of the data, but there are debunking approaches also are worth considering. Specifically, it may not be immediately obvious that the results cannot be explained just in terms of iconic memory or fragile visual short-term memory, or perhaps even regular working memory

Consider first the hypothesis that subjects' performance might be explained in terms of working memory. This could perhaps adequately explain subjects' performance in the trials where they are cued in advance: having been cued to look out for a picture matching the description 'wedding', for example, they might rapidly encode semantic information about each image as it is presented, specifically searching for a wedding and discarding information about all other images.

However, this account does not explain subjects' almost identical level of accuracy when they were cued *after* they had seen the images. As noted in the previous chapter, working memory has strict capacity limits, well below the twelve images presented in some trials. At most, then, it is likely that subjects would be able to retain just four images in working memory, fewer than the half the images in the array. That might still enable them to perform marginally above chance, but if that was indeed what was responsible for subjects' performance, then we would expect a significant difference in subjects' performance between the 6-item and 12-item trials, since they would go from being able to encode the majority of the stimuli to barely a third of them. In fact, subjects' performance was very similar in the two trials, suggesting the effect is not due to working memory alone.[12]

An alternative debunking hypothesis might claim that subjects were retaining a *sensory* representation of the stimuli, much as they do in the Sperling Test and other partial report paradigms. More specifically, one might imagine that subjects retain multiple distinct icons corresponding to each of the 6-12 images as they are presented, which they can then conceptualize and 'inspect' after presentation when they are given a target.

This sort of explanation seems unlikely, however. As noted earlier, visual forms of sensory memory like iconic memory and fragile visual short-term memory are both disrupted by the presentation of sequential images in the same location. This is not an issue for the Sperling experiment, since this involves just a single initial stimulus prior to cueing. In Potter's experiment, however, subjects saw *multiple* images one after another in exactly the same

[12] Potter et al. agree, commenting that "we can reject the hypothesis that participants could encode only two or three pictures... otherwise, performance would have fallen more dramatically in Experiment 2, especially in the after condition, in which participants had to retain information about the pictures for later retrieval" (Potter et al., 2014: 275).

location prior to cueing. Any representations in iconic or fragile visual short-term memory would therefore be rapidly 'overwritten' as the sequence was presented.

Additionally, a sensory interpretation would struggle to explain other experiments performed by Potter that examined subjects' vulnerability to conceptual 'decoys'. In one such experiment, subjects were shown a sequence of five pictures at 173ms exposures, and then immediately given a test picture and asked whether it was one of the five pictures just presented (Potter, Staub, & O'Connor, 2004). Subjects performed fairly well at this task. However, they were significantly more prone to error when tested on distractors which were similar in *semantic* content to pictures that had just been presented. These 'decoy' pictures were carefully chosen so as to share the same conceptual gist as one of the pictures shown by Potter without being too visually similar (see Fig. 2c).

This suggests that subjects' ability to remember the pictures was not based purely on the retention of low-level sensory information but involved encoding of the image in terms of its semantic properties. Otherwise, one would not expect the conceptual (but non-pictorial) similarity of the images to have any significant effect on subjects' performance. This is further evidence that subjects' performance in Potter's work on CSTM is not simply a matter of sensory short-term memory.

**Fig. 2c.** An example of a target picture ('camel') and its conceptual decoy. Figure from Potter et al. 2005.

Further insight into the experiments just described may be gleaned from a recent attempted replication of the earlier experiment of Potter et al. (2014) by Maguire & Howe (2016). Maguire & Howe used a variety of masking techniques to ensure that images in Potter's original study were not being stored in sensory buffers. The only masking technique used in the original study of Potter et al. were successive pictures of natural scenes. By contrast, Maguire & Howe used more aggressive masks (see Fig. 2d).

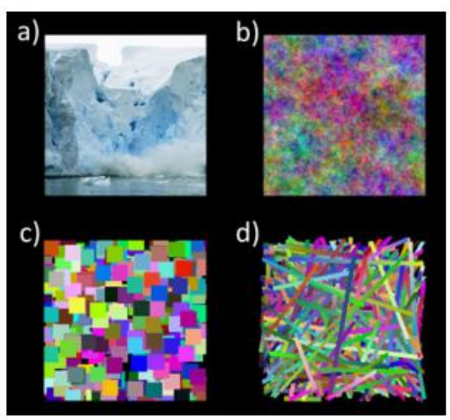



**Fig. 2d.** Masks used by Maguire & Howe. Mask (a) is a natural scene, Mask (b) is a 1/f noise mask, Mask (c) is a geometric mask, and Mask (d) is a colored lines mask. From Maguire & Howe, 2016.

Maguire & Howe's findings may seem initially to cast doubt on the CSTM hypothesis. Specifically, they found that for exposures of less than 53ms, the use of geometric and colored line masks caused subjects' performance to fall to near- or at-chance levels (though with 1/f noise masks subjects' performance was better than in trials with natural scenes masks). However, Maguire & Howe *also* found that subjects' performance for longer exposure times – 53ms or 80ms – was well above chance *regardless* of the mask used.

Maguire & Howe's data, then, suggests that masking can interfere with performance in trials like Potter's for short but not long exposures. This is not fatal for the CSTM hypothesis, however. Given that masking did not drastically hamper subjects' performance for longer exposures, it may simply be that Potter's results show two distinct processes at work, one (at the shorter exposures) involving sensory forms of memory like fVSTM and the other (at the longer exposures) involving storage in CSTM proper. However, this is just one way to interpret the evidence. It may also be possible that some form of sensory memory may be contributing to subjects' performance in trials involving short exposures. For example, encoding in CSTM may causally depend on prior strong activations in sensory areas which are in turn disrupted by short exposures combined with rapid masking.

Maguire and Howe's results, then, point to the need for further work, but are compatible with the CSTM hypothesis. In particular, the discovery that masking cannot eliminate subjects' performance for durations of 53ms or more if anything strengthens the evidence that there is a robust non-sensory form of memory at work, although given that Maguire and Howe used just pre-stimulus cues and 6-item sets of stimuli, we cannot rule out the possibility that this was simply working memory. Particularly useful for future work, then, would be a further experiment that combined Potter's original post-stimulus cueing and 12-item sets of stimuli with the more aggressive masking techniques of Maguire & Howe. If it was discovered that subjects still performed well above chance for 12-item sets even under conditions of post stimulus cueing and aggressive masking, this would constitute most the powerful evidence for CSTM yet.

## 2.4 – Evidence for CSTM II: Parallel Search

I now wish to present a different form of psychological evidence that has been advanced in support of the CSTM hypothesis. This comes from recent work on parallel visual search tasks, and seems to indicate that there is parallel processing of semantic information prior to activation in Central Cognition. Unlike Potter's experiments, most of which involve rapid serial visual presentation, these experiments involve presentation of a *single* array of stimuli, much like the Sperling experiment.

I will begin with a paradigm in which subjects were presented with an array containing multiple items (Moores, Laiti, & Chelazzi, 2003; see Fig. 2e).[13] Subjects looked at a central fixation point, and their task was to assess whether a target stimulus (for example, a motorbike) was present among the items in an array while their gaze direction was followed with an eye tracking camera. Moores et al. showed that the presence of semantically-related distractors in the array (e.g., a motorbike helmet) had a negative effect on subjects' reaction times and accuracy in trials where the target was absent. This suggests that subjects were distracted or otherwise impaired by the presence of pictures that were conceptually related to their target.

This is not itself surprising, and is compatible with an account cashed out purely in terms of working memory: it is possible, for example, that as subjects rapidly scanned the array they

[13] One might wonder why memory mechanisms are relevant to the following experiments, since the arrays in question remain in view of subjects at all times. However, as suggested in the introduction, any short-term processing of information, including perceptual encoding of semantic properties, presumes some kind of underlying short-term storage mechanisms, and I can no reason see why these mechanisms would be different in cases where the stimulus was still visible. Indeed, in one of the studies (Belke et al., 2008) the experimenters specifically take their data to support the CSTM hypothesis.

were momentarily sidetracked by the semantically related items, and had to pause to assess whether they fit the target description. What is much more surprising is that in a majority of trials where the target was absent but a semantically related distractor was present, subjects fixated the distractor *before* all other stimuli. In other words, subjects' looking behavior already seemed sensitive to the semantic properties of the items in the array even before they had the chance to visually fixate them.

This is difficult to explain just in terms of iconic memory, since, as noted, sensory forms of memory do not seem to be sensitive to high-level semantic properties of stimuli such as their specific category identity. However, in this trial, subjects' eye movements were sensitive to just such semantic properties of the items in the array.

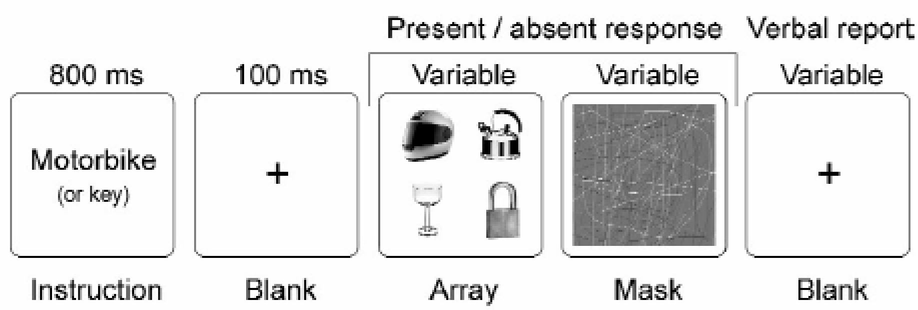



**Fig. 2e (from Moores et al. 2003**). Subjects are given a verbal cue, then fixate a central point. They are presented with an array of objects for brief durations (average 126ms) and asked whether the target stimulus is present. Eye-tracking showed that in a majority of trials subjects initially fixate the target object if present or semantically related distractors if the target is absent.

Nonetheless, since there are only four items in the array, this result could be explained in terms of working memory. For example, subjects might be engaging in a kind of very rapid covert attention to all the items in the array, rapidly extracting semantic information about all four of them to guide subsequent eye movements.

This interpretation looks much less plausible, however, in light of a later similar experiment (Belke et al., 2008). This experiment used a broadly similar methodology, but varied the number of objects in the trial, with arrays containing up to eight items. Crucially, this increase in the size of the array did not affect the likelihood that subjects initially fixated the target or semantically-related distractor object; instead, subjects immediately looked at the target or distractor object in a majority of trials.

This result clearly runs counter to the hypothesis that the experiments can be explained just in terms of working memory. If subjects were indeed just using working memory together with some form of covert attention to access the semantic properties of each item to begin with, then the size of the array should make a significant difference to how frequently and reliably subjects directed their initial eye movements to semantically relevant targets or distractors. Instead, subjects initially looked at semantically relevant objects with equal frequency in trials with 4 objects and trials with 8 objects. This suggests that the rapid semantic classification of objects in an array need not be limited by the resources of working memory.

Belke et al. also tested for the effect of cognitive load on subjects' performance. Specifically, subjects were required to remember a series of digits during the presentation of the array (see Fig. 2f). If subjects were using capacity-limited working memory to rapidly encode the semantic identity of each item in the array, one would expect their performance to be severely hindered by this task.

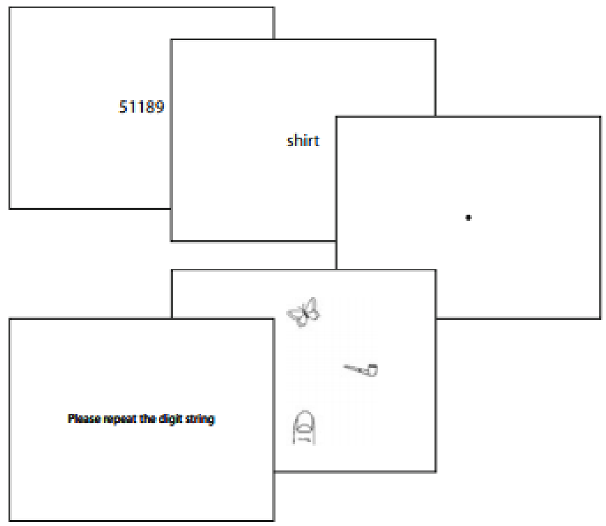



**Fig. 2f (Belke et al. 2008).** Subjects were briefly presented with five digits to hold in memory. They were then given a target description and told to search for an object matching that description in a presented array.

Instead, the results suggest a more complex picture. Although the cognitive load affected subjects' reaction times when *reporting* the presence or absence of a target, as well as broader features of their looking behavior (specifically causing subjects to linger longer on distractor items before moving on), it did not affect subjects' initial direction of gaze at all: they were just as likely to direct initial visual attention to target and distractor items in cognitive load cases as in all other variations of the experiment. This suggests that the mechanism by which the semantic identity of visual stimuli is retrieved (and which directs initial eye movements) does not rely on the same cognitive resources as Central Cognition. As Belke et al. observe, their findings support the CSTM hypothesis, and further, extend it, by "suggesting that there is parallel conceptual processing of visual stimuli prior to the first selection for attention" (Belke et al., 2008).[14]

Although I take the evidence from Belke et al. to provide strong support for Potter's

[14] The evidence for CSTM described above suggest a further puzzle for interpretation of earlier work on iconic memory: why, if there is rapid conceptual processing of items presented in an array, did subjects in Experiment 6 is Sperling's original paper fail to exhibit partial report superiority when cued to report on items in respect of their *semantic* identity (e.g., "letter" or "number")? There are several interesting possible explanations for this, but I will mention one possibility here, namely that subjects may have encoded the items in the array in respect of more specific semantic categories ("r" or "7") than those cued by Sperling ("letter" and "number"). Thus the number of characters that were reportable would be limited by subjects' ability to make rapid inferences about the determinable class the individual items belonged to. This would be a valuable area for further experimental work.

CSTM hypothesis, it is worth considering some debunking hypotheses. As noted above, it does not seem likely that subjects' performance can be explained just in terms of working memory, given that the size of the array did not make a difference to their eye movements. It is also hard to see how it could be explained in terms of sensory forms of memory, given that subjects' gaze fixation seemed sensitive to the semantic connections between the target item and distractors.

However, there is one proposal that is worth considering. Specifically, it seems plausible that when subjects are cued in advance to look for a given conceptual target (e.g., "motorbike"), they thereby activate some stored visual template in long-term memory to aid their search. The activation of this template may consequently (perhaps through some process of spreading activations) cause other visual templates to become active, thereby making subjects more likely to direct their eye movements towards targets that match these templates.[15] If this view was correct, then subjects' performance could be explained just in terms of the activation of associated sensory templates without subjects processing semantic information about most of the items in the array.

This is a reasonable proposal, and currently the science does not allow us to decisively say whether it is better than the CSTM explanation endorsed by Belke et al. However, one piece of evidence that might be taken to support the CSTM explanation comes from similar work on parallel visual search in real world scenes by Hwang et al. (2011). This paradigm, like that of Belke et al., demonstrated that subjects' eye movements are heavily influenced by the semantic properties of the objects in the scene, and specifically their eyes movements are likely "to transition to objects that are semantically similar to the currently fixated one" (Hwang et al., 2011). Crucially, they found that this tendency was *not* influenced by the length of the saccade involved; in other words, subjects' gaze tended to flit between semantically-related objects to a

[15] Thanks to Will Bridewell for this suggestion.

similar extent even when this involved long saccades far from the initial point of fixation.

This may be important because of the relatively poor resolution of parafoveal and peripheral vision: if subjects were relying on the matching of sensory templates to plan and execute eye-movements, then one might naturally expect this process to be less robust in cases where the target was perceived with a low resolution (i.e., in peripheral vision) such that their eye movements would less reliably transition to semantically-related items in their parafoveal and peripheral vision. However, as noted, this is not what happened. By contrast, the CSTM hypothesis can more readily explain the results. Specifically, if subjects' looking behavior is controlled primarily by the semantic properties of items rather than the matching of sensory templates, the diminished resolution of peripheral vision would not decrease the likelihood of subjects saccading between semantically-related items, as long as the visual resolution of items in the periphery was minimally *sufficient* to enable the visual system to extract basic information about the semantic category of the objects in question. In other words, if it is the representation of semantic categories rather than the matching of sensory templates that is responsible for directing subjects' gaze, then the fact that some objects are encoded at a lower spatial resolution than others would not make them less likely to be targets, provided that this lower resolution still enabled their semantic classification.

This is just one piece of evidence in support of the CSTM interpretation of Belke's results, and is not meant to be decisive. Moreover, it is worth noting that the role of semantic information in guiding eye-movements and attention is still hotly debated, and the consideration just mentioned may be challenged by future science. However, as matters stand, Belke et al.'s seem to have good grounds for their claim that the best explanation of their results is the CSTM hypothesis, that is, that there exists a stage of processing in which the outputs of rapid parallel semantic processing are made briefly available.[16]

## 2.5 – Evidence for CSTM III: the Attentional Blink

A third and final piece of evidence for CSTM comes from work on the attentional blink phenomenon. This is a strange effect whereby subjects in rapid serial visual presentation tasks can fail to notice a normally visible object if it is presented in sequence *after* a specific target. Specifically, subjects typically not report seeing a test item if it is presented 200-500ms *after* a target item (see, for example, Sergent, Baillet, & Dehaene, 2005). However, if subjects are given the same experimental setup without being assigned a task, they notice all items without difficulty. A similar effect can be obtained by presenting subjects with a task-irrelevant word immediately *prior* to a target: in this case, subjects report not having seen the task-irrelevant word (Kouider et al., 2007). To summarize these paradigms, it is as if the cognitive effort involved in recognizing the target item and holding it in mind causes a brief dip in subjects' ability to spot subsequent objects (and causes them to forget having seen items presented immediately before).

The attentional blink offers a unique way of suppressing subjects' ability to report on items with minimal impairment of overall sensory processing: it does not involve degraded or masked stimuli (as in many paradigms involving unconscious perception), and it does not rely on diverting subjects' attention away from a stimulus at the very same moment that stimulus is

[16] A final puzzle arises, however, if we endorse the CSTM explanation of the above results which relates to the Sperling Task. Why, if there is rapid conceptual processing of items presented in an array, did subjects in Experiment 6 of Sperling's original paper fail to exhibit partial report superiority when cued to report on items in respect of their semantic identity (e.g., "letter" or "number")? There are several interesting possible explanations for this, but I will mention one possibility here, namely that subjects may have encoded the items in the array in respect of more specific semantic categories ("r" or "7") than those cued by Sperling ("letter" and "number"). Thus the number of characters that were reportable would be limited by subjects' ability to make rapid inferences about the determinable class the individual items belonged to. This would be a valuable area for further experimental work.

presented (as in inattentional blindness paradigms). Instead, a stimulus that is normally visible is presented immediately in front of a subjects' eyes at a moment when they are required to do nothing else; and yet subjects report failing to see it.

It is an open question whether attentional blink experiments involve the absence of conscious perception, or merely of reportability (see Chapter 4). However, what is most important for present purposes is that there is strong psychological and neuroimaging data that suggests that that subjects in attentional blink experiments engage in short-term *semantic* processing immediately after presentation of a stimulus, even if they do not access in it Central Cognition. For example, in one study by Luck, Vogel, & Shapiro (1996), subjects saw a sequence of letters, numbers, and words for 83ms each, and were required to report whether a target string of numbers began with an odd or even number. At one point in the sequence, a probe word was shown, which was either semantically related or semantically unrelated to an initial anchoring word presented at the start of the sequence (see Fig. 2g).

| Stimulus Type | Time (ms) | Related Trial | Unrelated Trial |
|---|---|---|---|
| Context Word | 1000 | RAZOR | WHEEL |
| Blank | 1000 | | |
| Distractor | 83 | PNVCSZP | KDSWPVZ |
| Distractor | 83 | GRSDPKN | VNMCPKL |
| Distractor | 83 | BVCPLMS | FDPMCNV |
| Distractor | 83 | DSPWTFR | VPMTDZM |
| Distractor | 83 | RLDJHGK | HJDLGFP |
| Distractor | 83 | SPLDJMF | DFPLJKH |
| First Target | 83 | 3333333 | 4444444 |
| Distractor | 83 | WDPTBNF | GHJDMVT |
| Distractor | 83 | SCDPVBF | HDVCBNM |
| Probe | 83 | XSHAVEX | XJEWELX |
| Distractor | 83 | FDLNLKB | NMCVPHJ |
| Distractor | 83 | DLJJCNW | DCVPBJM |
| Distractor | 83 | WPSCDSN | PCNBVLK |
| Distractor | 83 | DPWVCPB | NPMTVDK |
| Distractor | 83 | CBNDPNJ | BRTFPMF |
| Distractor | 83 | RTPMVBC | JLSDCDK |
| Distractor | 83 | TWSCLMN | LKSDVCP |
| Distractor | 83 | LJVBCMH | DKKHNVP |
| Distractor | 83 | RMVCPKL | WKLDMZP |
| Distractor | 83 | DPNMNVZ | CPNHVGB |
| Blank | 1000 | | |
| Response Cue | 2000 | ? | ? |
| Blank | 2000 | | |

Fig. 2g. Subjects saw a task-irrelevant context word and were told to identify whether a sequence of numbers (labelled First Target) began with an even or odd number. A probe word was then display shortly after presentation of the number which could be either semantically related or unrelated to the context word. In semantic mismatch trials, subjects exhibited a large N400 even though they were largely unable to subsequently report (1-2 seconds later) the probe word in question. This suggests semantic processing can occur if subjects are unable to report on an item's identity or are unaware of its having been presented. From Luck, Vogel, & Shapiro (1996)

As expected in an attentional blink paradigm, subjects frequently failed to notice the probe word. Fascinatingly, however, subjects exhibited a strong N400 signal in all cases where the probe word was semantically unrelated to the anchoring context word. The N400 signal is widely found in any case of semantic mismatch between words and categories. The presence of the N400 just in trials where there was a mismatch between the anchor word and probe indicates that subjects at some level accessed the semantic content of both words. However, the fact that this occurred even on trials where subjects claimed not to have seen the words in question suggests that there was no encoding in working memory. The authors of the paper find this puzzling, and note that "it is difficult to determine whether the probe words were identified without reaching awareness or if they momentarily reached awareness and were then rapidly forgotten." Subsequent follow experiments convinced Vogel & Luck in particular that there is some form of short-term semantic processing prior to Central Cognition (Vogel, Luck, & Shapiro, 1998), and they appeal to Potter's CSTM hypothesis to explain this.

> Following Potter (1993), we begin by proposing that all items in the RSVP stream are initially stored in a conceptual short-term memory (CSTM) buffer after being fully identified (i.e., after reaching a conceptual level of representation). At this stage, the items are not yet available for report and are prone to decay and to replacement by other incoming stimuli (see Enns & Di Lollo, 1997). We further propose that attention serves to consolidate information stored in the CSTM buffer into a reportable and more durable form, which we call visual working memory.

To summarize, then, we have rich data from multiple paradigms that point to the existence of a mechanism for short-term semantic processing of visually presented information outside of capacity-limited Central Cognition, where this is not vulnerable to disruption by new visual input. The CSTM hypothesis explains this data by positing a high-capacity, fragile, short-term conceptual store. While many questions remain concerning CSTM – for example, whether the same mechanism is at work in different sensory modalities – I contend that we already have good reason for taking CSTM seriously.

## 2.6 – CSTM and the Revised Model

If data presented thus far withstands further experimental scrutiny, then I would suggest we have reason to think there are forms of short-term processing that cannot be readily accommodated within the Sensory-Cognitive model presented in the previous chapter. Consequently, we may need to emend the model so as to include CSTM or some closely related faculty. My next goal for this chapter, then, will be to suggest a way in which we could revise the Sensory-Cognitive model to accommodate CSTM, as well as briefly sketching the role CSTM may play in perception and cognition.

The central hypothesis about CSTM that I now wish to explore can be briefly stated in two main claims. The first claim is simply that that there is at least one mechanism supporting brief semantic processing and storage of perceptual information prior to the engagement of capacity-limited Central Cognition. This I take to be supported by (i) the findings of Potter et al. (2014) and Belke et al. (2008) that subjects display a sensitivity to the semantic properties of presented items that is preserved even when the number of objects exceeds the capacity of Central Cognition; and (ii) the findings from work on the attentional blink (Luck, Vogel, & Shapiro, 1996; Vogel, Luck, & Shapiro, 1998) that show neural detection of semantic mismatch even in cases where subjects report not having seen a stimulus (and therefore presumably did not encode the stimulus in Central Cognition).

The second claim is that this semantic processing is not merely a sensory process; in other words, that it is not merely a form of late modality-specific nonconceptual processing. My evidence for this claim comes from the fact that all previous work on sensory memory has indicated that it is vulnerable to overwriting by serial presentation (Sligte, Scholte, & Lamme, 2008; Averbach & Coriell, 1961; Saults & Cowan, 2007, show overwriting effects in echoic memory), whereas the semantic processing discovered by Potter et al. (2014) was not vulnerable to these effects, at least for exposures greater than 50ms (Maguire & Howe, 2016). That alone suggests that none of the currently posited sensory memory mechanisms (such as iconic memory and fragile visual short-term memory) can explain the data.

If these claims are correct, then they suggest that we should emend the Sensory-Cognitive Model to include a form of semantic memory intermediate between sensory member and Central Cognition. This leaves us with the following picture (Fig. 2h).

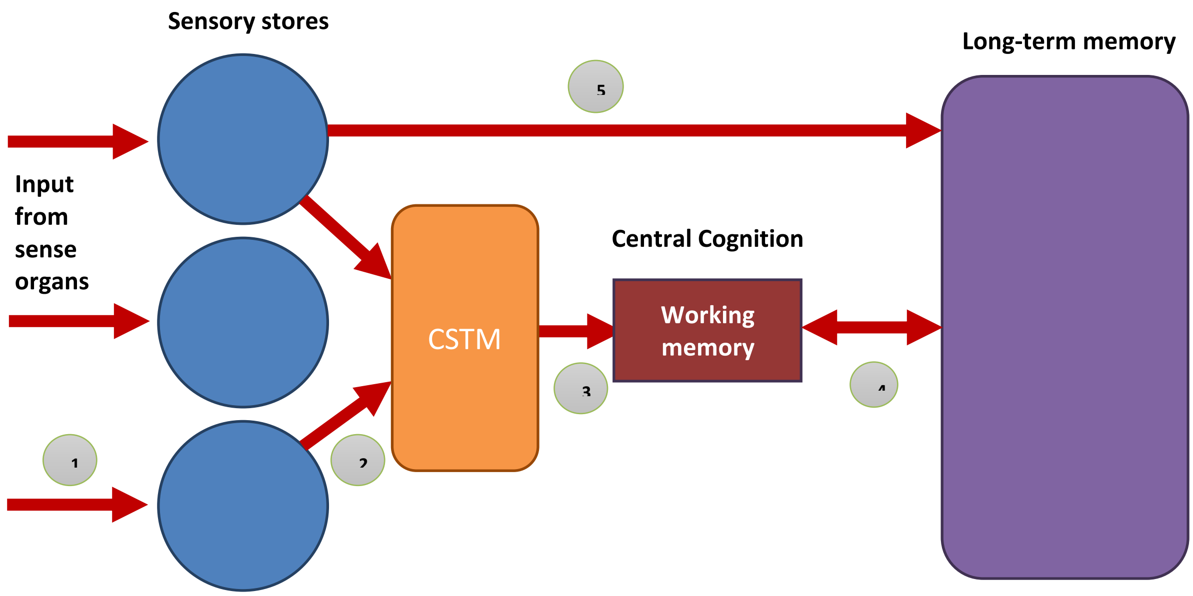



Fig. 2h. The Revised Sensory-Cognitive Model showing the role of CSTM.

1. Encoding of inputs in sensory memory. 2. Access and semantic encoding of sensory memory by CSTM. 3. Selective access of CSTM by working memory. 4. Reading from/writing to LTM by Central Cognition. 5. Sensory information triggers activations in LTM by association without cognitive access?

As the figure shows, CSTM takes its inputs from sensory stores such as iconic memory and makes them available to Central Cognition. Items stored in Central Cognition can then be encoded in LTM. Each of the four stages amounts to a distinct kind of processing which can be (somewhat cautiously) characterized in terms of its psychological role and characteristic ERPs (see Fig. 2i below).

If this model is correct, it amounts to a serious emendation of the standard picture. My broader hypothesis, to be explored in the following chapters, is that the basic function of CSTM is rapid and effortless perceptual classification in respect of semantic categories. This may range from simple low level classification of, say, a shape as being a square, but may include, as Potter shows, high level concepts like "wedding" or "picnic'.

| **Memory store** | **Capacity** | **Duration** | **Neural signatures** |
|---|---|---|---|
| Sensory stores | Large; varies | Up to 1s (iconic memory); 4s (fVSTM). | P1, C1 |
| CSTM | At least 8 items. | 1-2 seconds? | N400 |
| Central Cognition | 4 items | Minutes | P300 |

**Fig. 2i. A comparison of the major memory stores.**

However, there are also a number of respects in which this broader model remains incomplete, reflecting the still patchy data pertaining to CSTM. Even within the model as described, then, we are left with several open questions, as I will now describe. Depending on the fate of CSTM, then, these areas may be fruitful targets for future research.

*(i) Can Central Cognition influence semantic processing in CSTM?*

It is unclear whether central cognitive processes such as storage of information in working memory (for example, the name of a target) can modulate processing of semantic information in CSTM. While results of Belke et al. may suggest that top-down attention can be attracted to targets in a way that is sensitive to their encoded properties in CSTM, it does not settle whether the kind of semantic information that gets processed by CSTM *in the first place* can be influenced by Central Cognition. Thus imagine that I am shown a series of words in various colors and told to look for a red word that is related to some target concept (e.g., "fruit"). Clearly, I can selectively direct top-down attention just to red words. However, it is unclear whether the CSTM can prioritize processing of sensory information to reflect Central Cognitive processes, such that, for example, the semantic identity of words in red is encoded in CSTM in preference to the semantic identity of non-red words.

*(ii) Direct encoding from CSTM to LTM.*

Another unanswered question concerns whether all encoding to long-term memory is mediated by Central Cognition, or whether information briefly stored in CSTM can give rise to representations in long-term memory without the involvement of Central Cognition. In other words, could it be that that we can store briefly glimpsed semantic information for longer durations, even if it is never accessed by Central Cognition? There is some evidence that this is *not* the case. Endress & Potter (2014) showed subjects numbers of images (more than 100) very quickly at a rate of 8/s. When given subsequent recognition trials half an hour later, subjects performed almost at chance. The authors concluded that "when items are viewed briefly, the resulting memories are fleeting and disappear over the course of a few minutes". This may suggest, however, that Long-Term Memory itself may not be a monolithic concept, but exhibit forms that last for shorter or longer durations. This a proposal developed in, for example, Ericsson and Kintsch's (1995) notion of Long-Term Working Memory.

*(iii) Encoding in Central Cognition without CSTM*

A final question concerns whether CSTM is a true gatekeeper for Central Cognition, or whether encoding of sensory items can bypass CSTM all together. One small piece of evidence comes from the fact that patients with associative agnosia lose the ability to recognize objects but not to copy images or make discriminations based on low-level properties like color (Greene, 2005; see Appendix 5). If CSTM functions to enable recognition of objects' categories, it is possible that associative agnosics have suffered severe impairments to CSTM. It might then be the case that agnosics' ability to copy images involved direct access of sensory representations by Central Cognition. However, another hypothesis would be that information is still routed through the (severely impaired) mechanisms of CSTM on its way to Central Cognition.

*(iv) Does CSTM operate across different modalities?*

A final more complex question that is worth briefly dwelling on concerns whether CSTM operates across multiple sensory modalities. I take it that the apparent sensitivity of CSTM to high-level semantic properties such as word meanings suggests that it does not encode content solely in terms of modality-specific features such as visible color, shape, size, and so on. However, this does not rule out the possibility that CSTM is a semantic buffer tied specifically for *visual* perception, leaving open the possibility that other senses may or may not possess equivalent forms of short-term semantic memory.

Given our seemingly effortless ability to rapidly classify speech and other sources of noise, audition might be a particularly promising place to look for rapid high-capacity semantic processing of the kind that I have suggested CSTM performs for vision (although care must obviously be taken to control for the contribution of sensory stores, and in particular the relatively longer duration of echoic as compared to iconic memory).

One intriguing experiment in this regard is the discovery by Daltrozzo et al. (2011) of semantic priming for auditorily-presented words just below hearing threshold. Daltrozzo et al. showed that a prime word was able to influence subjects' performance on lexical decision tasks for semantically related words even when subjects were not able to classify the category of the prime word. Some subjects *did* reliably report having heard some word when the prime was presented, but claimed not to know its semantic category, despite this influencing their behavior in the subsequent tasks. This might be taken to suggest that there was semantic processing in the absence of access by Central Cognition. Additionally, Daltrozzo et al. showed that the effect on performance in lexical decision tasks was relatively greater for words semantically related to the prime word as compared to trials in which the prime itself was used as the target. This suggests that this was not simply a very low-level effect due to the activation of some sensory template combined with activations that spread to semantically-related words, since such an account would presumably predict greater facilitation on lexical decision tasks when the prime itself was reused (since its sensory template would be identical).

This result might be very tentatively taken as evidence for something like CSTM in the auditory domain. However, further work would be required to rule out other hypotheses. Moreover, it says nothing about whether the mechanism in audition is the same mechanism as seems to be present in vision or is instead tied to audition specifically. However, this result provides yet more promising evidence for a stage of rapid semantic processing outside Central Cognition, adding to the already rich body of data presented thus far.

## 2.7 – CSTM and Unconscious Processing

In the next two chapters, I will go on to explore how the CSTM hypothesis may relate to philosophical work on perception and consciousness, arguing that it may have an important contribution to make to these debates. Before turning to these more properly philosophical considerations, however, it is worth briefly surveying how the CSTM hypothesis relates to research on unconscious semantic processing. The CSTM hypothesis, after all, is framed around the claim that there is a specific mechanism for non-sensory processing of semantic information outside of Central Cognition, and there are a number of other experiments that have purported to show that semantic processing can occur for items without access by Central Cognition and seemingly without awareness (see Dehaene, 2014, Ch.2, for a review). For example, Dehaene et al. (1998) demonstrated that subliminally presented numerals could facilitate or impair subjects' subsequent performance on simple mathematical tasks, and a study by Naccache et al. (2005) showed that subliminally presented happy or sad faces activated areas of the brain associated with the processing of emotions. Could paradigms such as these perhaps be explained in terms of CSTM?

While it is possible that CSTM may have a role to play in explaining the rich data on unconscious semantic processing, I wish to resist the conclusion that we should confidently appeal to it to explain these results, for several reasons. The first reason is that I want to remain open to the possibility that information stored in CSTM is frequently or even always conscious. If this is correct, then insofar as there are clear cases of unconscious semantic processing these may involve quite different mechanisms. The relationship between CSTM and consciousness will be the focus of Chapter 4, so I will not defend it in detail here. However, it is worth noting that Potter herself is at least open to the possibility, stating that "the evidence… demonstrates that there is conceptual processing of material that is subsequently forgotten, [but] it does not tell us whether we were briefly conscious of that material, or whether the activation and selection occurred unconsciously." (Potter, 2012: 8).

A second reason I think we should be cautious about invoking CSTM to explain the broader data on unconscious semantic processing is that the experimental procedures appealed to in support of CSTM have a number of important differences from the paradigms typically used to test for unconscious semantic processing. For example, the work of both Potter and Belke et al. involves stimuli that are fully visible to subjects, and which can be reported given suitable cues. By contrast, most paradigms that claim to show unconscious semantic processing involve stimuli that a subject is wholly unaware of and cannot be accessed even when cued (see Kouider & Dehaene, 2007). Instead, evidence for unconscious semantic processing in these cases comes from the appropriate facilitation by these subliminal stimuli on subsequent tasks. Admittedly, subjects in the attentional blink trials discussed in 2.5 also claim not to have seen the target item. However, in these trials the stimuli themselves are not masked or suppressed in any way, and are fully visible under control conditions when the subjects is not engaged with other tasks, leaving open the possibility that they are consciously experienced (see, e.g., Phillips 2015). By contrast, the stimuli used in the clearest cases of unconscious semantic processing are not reportable under any conditions.

A final point worth bearing in mind is that unconscious perception, including unconscious semantic processing, may involve a variety of different mechanisms (Block, 2011b). This is suggested by the quite different time courses observed in different experimental paradigms. Thus whereas semantic processing of subliminally presented numbers happens fairly quickly (and can be detected using ERP measures after 300ms; Dehaene et al., 1998), unconscious processing of emotional stimuli seems to happen much more slowly (with differences detectable using ERP measures only after 800ms; Naccache et al., 2005). This points to the possibility that unconscious semantic processing may in fact take place across multiple different pathways depending on, for example, the content of the stimuli and the kind of masking involved. If this is right, then some or all of this processing may take place without the involvement of CSTM.

For now, then, I would suggest that CSTM should be seen as one possible missing mechanism in a still incomplete picture of the broader architecture of the mind, while remaining open to the possibility that are other kinds of semantic processing that can take place unconsciously or outside of Central Cognition. Evidence of high-capacity, brief duration, and the encoding of semantic content can help us identify whether CSTM is active in any given case, but ultimately, I am inclined to see CSTM as a distinctive psychological kind, thereby remaining open to the possibility that there may be other mechanisms that are (at least in some ways) functionally similar to CSTM.

## 2.8 – Conclusion

This chapter has surveyed a wide range of evidence for a new form of memory, namely CSTM. I have presented a range of data from RSVP, parallel search, and the attentional blink that arguably all point to the presence of a distinctive stage of processing that has been left out of the Sensory-Cognitive Model. I have also suggested that the defining features of this mechanism are its high capacity, brief duration, immunity to disruption by pattern masks, and its encoding of inputs in respect of semantic contents. The revised picture of short-term memory presented in 2.5 above, I suggest, may offer a distinct improvement on the Sensory-Cognitive Model, though one that is of course open to refutation by future research. I have also argued that we should not automatically assume that CSTM is responsible for all semantic processing outside Central Cognition, allowing that there may be multiple mechanisms involved across different experimental paradigms.

In proceeding, I wish to stress again that the ultimate fate of most of the claims made in this chapter will depend on future psychological work. For example, it may ultimately emerge that the effects attributed to CSTM in this chapter can be explained by new Variable Encoding models of working memory (see Appendix 2) or by theories of Unconscious Working Memory (Appendix 3). However, it is far from clear that such approaches as they currently stand could accommodate the range of data presented thus far.

Moreover, even if some of the specific hypotheses about CSTM considered in this chapter turn out to be false, we may still find that some closely related emendations to the Sensory-Cognitive model are required to accommodate, for example, *some* kind of larger-capacity stage of processing outside of conscious working memory. Thus I hope that the broader philosophical claims and ideas of this dissertation will be of independent interest even if evidence mounts against CSTM, although they will of course be of much *greater* interest if further data mounts in support of it.

I will proceed, then, on the working hypothesis that CSTM is a distinctive part of the mind, and turn from specific empirical debates to the exploration of broader philosophical ones. My immediate goal in the next chapter will be to argue that CSTM may underlie what is known as *categorical* perception, and in this regard, may make a distinctive contribution to perceptual experience. I will then go on in Chapter 4 to consider whether CSTM may also have a role to play in a broader theory of the psychological basis of consciousness.

# CHAPTER 3: CSTM at the Border of Perception and Cognition



## 3.1 – Introduction

In the previous chapter, I presented evidence to the effect that the Sensory-Cognitive model is incomplete – a part of the mind is missing. Specifically, I suggested that what is missing is CSTM, a short-term memory buffer responsible for the rapid encoding of perceptual representations in respect of their semantic properties.

If this proposal is even broadly along the right lines, I believe it may have significant ramifications for a variety of debates in the philosophy of mind. In this chapter, I wish to examine one set of debates where CSTM may be of use, all related to the perception-cognition distinction. Specifically, I wish to argue that once we admit of a stage of semantic short-term memory processing intermediate between sensory processes and Central Cognition, we may be better placed to account for a ubiquitous and puzzling phenomena, namely *categorical perception*. Working on the assumption that categorical perception is accomplished by CSTM, we can then offer new perspectives on debates about high-level phenomenal properties, and certain putative cases of the cognitive penetration of perception.

I believe that my proposals in this chapter are of independent interest for philosophers invested in these debates. However, to the extent that my claims are compelling, it will also serve to somewhat demystify the role of CSTM in perceptual experience, and perhaps lend independent philosophical weight to the thus-far largely empirical claim that the Sensory-Cognitive model is missing something out.

The chapter will proceed as follows. First, in 3.2, I will summarize the problem of categorical perception, and briefly outline how CSTM may contribute to resolving it. Second, in 3.3, I will show why it is difficult to account for categorical perception in strictly perceptual or strictly cognitive terms. In 3.4, I will expand upon the suggestions given in 3.2, and give a more developed account of the role of CSTM as the constitutive basis of categorical perception. I go on in 3.5 to examine how this model may apply to debates about high-level phenomenology. Finally, in 3.6, I speculate as to how the other proposals in this chapter may be of relevance to debates about the alleged cognitive penetration of perception.

## 3.2 – The problem of categorical perception

In Chapter 1, I suggested that a number of authors draw the perception-cognition distinction in a way that can naturally be accounted for in the Sensory-Cognitive model. Specifically, authors like Burge, Block, and Dretske all claim that perception, unlike cognition, has a large capacity and a distinctive sensory format. Moreover, they endorse the idea that Sperling's iconic memory provides evidence of a distinctively perceptual stage of mental processing, while pointing to work on limited capacity working memory as evidence of a set of central cognitive mechanisms underlying conscious thought.[17]

Many philosophers would undoubtedly challenge the idea that we can distinguish perception and cognition by virtue of format and capacity; indeed, some would deny that we can draw this distinction at all.[18] However, my arguments in this chapter will proceed within the broad set of frameworks that accept this way of drawing the perception-cognition boundary. My goal in doing so will be to show that, framed in these terms, the perception-cognition distinction struggles to accommodate the phenomenon of categorical perception, and in turn that we can offer a better explanation of the data by expanding the model to include CSTM (or some similar faculty of semantic processing intermediate between perception and cognition).

To begin with, then, I will spell out what I mean by categorical perception, or as it is sometimes known, 'perceiving-as'. To illustrate the phenomenon, it is worth distinguishing two distinct aspects of perceptual experience. On the one hand, we can characterize your perceptual experience in terms of how it makes you aware of objects in respect of their shape, size, and color. Thus, as Prinz puts it, "[i]n vision, we experience stimuli as bounded wholes from a specific vantage point, occupy a specific size and position within the visual field" (Prinz, 2011b). However, visual experience also serves to make us aware of the semantic *identity* of objects. Thus imagine that you are walking down the street when you see your old friend James. Your visual experience indeed serves to make you aware of James as a bounded whole of a certain size standing in a particular perspectival relationship to you. However, it arguably also serves to make you aware of him *as a human being* and *as James.*

[17] Indeed, Block goes further, suggesting that conscious perception, at least, is in fact *constituted* by the storage of information in fVSTM (Block, 2007).

[18] See, for example, Andy Clark, who claims that "the lines between perception and cognition [are] fuzzy, perhaps even vanishing'' (2013: 190). Similarly, Gary Lupyan (2015) states that he supports "a collapse of perception and cognition which makes the whole question of the penetrability of one by the other, ill-posed."

This latter form of awareness is what is commonly termed categorical perception or perceiving-as. It can be at least conceptually distinguished from perception in the narrow sense of the term insofar as it draws upon one's background knowledge (of humans and of James, for example) and involves the application of particular categories to what you are seeing. Perception, by contrast, at least by the lights of many theorists (including Prinz, 2012, and, with some exceptions, Burge 2010a), encodes objects just in respect of more basic properties like those just mentioned.[19] It can also again be at least in principal distinguished from simply forming thoughts about what we see. Thus, looking at my friend James I may judge that he has taken the day of work, and that he is probably on his way to the supermarket. However, it does not seem appropriate to say that I *see* him as having these properties, at least in the same sense that I see him as being James.

Before proceeding, I wish to briefly flag two questions about categorical perception that I wish to set aside from the time being. The first question is whether categorical perception has its own distinctive phenomenology. One way to present this question is to ask whether, for example, seeing James *as* James *directly* alters the character of my experience, in a way that is not just a matter of, for example, attending to certain features (his distinctive moustache, his cleft chin) rather than others. This is matter of hot dispute, and I will return to it later in the chapter.

A second important question concerning categorical perception is whether *all* perception has a categorical element – that is, whether it is ever possible to see something

[19] Note that I will not give a worked out account of how to distinguish whether a given content is high-level or low-level, instead relying on our intuitive understanding of the notions and clear exemplars; thus, I take it that representations of shape and color are fairly low-level properties, while representations of properties like 'being Donald Trump' or 'being a CD player' are high-level. There is, to my knowledge, no good consensus on how to draw this distinction more precisely, and there may be no truly rigorous way of doing so. See Logue (2013) for a good discussion of this question.

without seeing it *as* anything in particular. Again, this is a matter of intense controversy: whereas Dretske (1981) claims that one can perceive something without thereby perceiving it as anything in particular, Fodor insists that "there is no perceiving without perceiving as" (Fodor, 2015). Once more, I will set aside this issue, returning to it in Chapter 4.

With these considerations out of the way, I now turn to my central question, which is where we should locate categorical perception within the operation of the mind. On the one hand, it seems to involve the use of learned high-level categories or concepts – attributives like 'James' or 'human', which might tell in favor of taking it to be a species of thought or judgment. On the other, though, it bears some of the hallmarks of perception. For example, it is fast and involuntary: I just have to open my eyes to see James as James. Likewise, it does not seem to occur in the total absence of lower-level properly perceptual content: in every case where I see James as James, I also undergo perceptual experiences involving representation of shape, color, distance, and so on.

This odd mix of features has led some theorists such as Bayne & Montague (2011) to wonder whether "some instances of 'perceiving-as' ought not be regarded as purely perceptual but are intermediate (or perhaps 'hybrid') states that straddle the divide between thought and perception." I note these features in passing, and will consider them more below. However, as a starting point, it may seem as though this is a place where a CSTM theory could do useful work. I suggested in the previous chapter that CSTM is a dedicated mechanism that serves to encode perceptual inputs in respect of some higher-level category, and this seems to describe exactly the function of categorical perception. Applying this to the expanded version of the Sensory-Cognitive model presented in the last chapter, I wish to explore the idea that we might distinguish perception proper from categorical perception by appealing to the distinction between sensory processes and CSTM. If this is correct, then part of the reason that categorical perception has puzzled philosophers and psychologists is that it relies on a piece of the mind that has hitherto been overlooked, and that by studying CSTM, we may be able to getting a better grasp on the phenomenon of categorical perception itself.

I will shortly say a little more about how this might work. However, I think much of the motivation for this view will come from examining in more detail the challenges involved in identifying categorical perception with perception or cognition in their strict senses. Therefore, I'll now first consider and cast doubt on the idea that categorical perception might be accomplished by the same mechanisms as perception proper, before considering and spelling out problems for the competing claim that it is just a matter of forming central cognitive beliefs. With these problems made explicit, I will then return to the idea that categorical perception might be best understood in terms of the function of some intermediate mechanism such as CSTM.

## 3.3 – Categorical perception: perception or belief?

### *(i) Is categorical perception just perception?*

Consider first then the idea that categorical perception might be accomplished by the same mechanism as perception proper; that is (within the present framework), a strictly *sensory* mechanism. To begin with, I will consider a specific proposal to the effect that many cases of 'seeing-as' are strictly perceptual processes, that of Ned Block (2014). Block's case stems from a series of experiment on perceptual adaptation. Simplifying somewhat, adaptation is a phenomenon whereby presentation of a given stimulus with some property *F* makes a subject less likely to see a subsequent stimulus as being *F*. In other words, perceptual adaptation occurs when perceiving a stimulus of a given type *raises the threshold* for seeing subsequent stimuli as being of the same type. It thus constitutes, as Block notes, the opposite phenomenon from priming.

I will not consider the full range of Block's cases here, but will instead focus on a single paradigm that will be illustrative of the central proposal of the paper. This was an experiment by Butler et al. (2008) that showed that adaptation occurs for seeing a face as angry or scared. Specifically, the study showed that subjects who fixated an angry face for 5 seconds were more likely to judge an ambivalent facial expression (neutral between fearful and angry) displayed for 300ms as being afraid, with the effect reversed when the initial face was afraid (see Fig.3a) .

**Fig. 3a –** Angry, neutral, and scared faces. From Block (2014), adapting Butler et al. (2008).

This data suggests that there is adaptation to emotion in the perception of faces. While, in this particular experiment, follow up trials suggested that the adaptation operated at the level of individual characteristically angry or fearful *features* rather than the whole face, other data presented by Block (Susilo et al., 2010) shows a form of adaptation to faces that is highly sensitive to orientation, suggesting that at least some adaptation operates at the level of whole faces rather than individual features.

Block concludes on the basis of trials like this that we have good evidence for the existence of genuinely perceptual (that is, sensory and non-conceptual) adaptation to some fairly high-level properties such as emotion and facial expression. If this is correct, he argues, we have good reason for thinking that processes involved in seeing a face *as angry* may be carried out at the perceptual level.

There are several points I wish note about Block's argument. First, even if adaptation can be used to ground some cases of categorical perception as being truly perceptual, it seems unlikely to explain very high-level categorical perception such as seeing an individual as Donald Trump, or seeing a an object as a stapler. Indeed, both Block (2014a) and Burge (2014a) are skeptical that such properties are in fact present in perceptual experience at any level below that of propositional judgment.

Second, I am hesitant about accepting Block's interpretation of the data without qualification. In particular, I would suggest that the evidence does not rule out the possibility that the representation of faces as angry or fearful was *conceptual* rather than sensory. Block gives two reasons for thinking this is not the case. He first notes that there seems to be no evidence thus far of strictly cognitive non-perceptual forms of adaptation, and observes that "if there were conceptual adaptation one would expect conceptual-without-perceptual adaptation." However, even if we grant the hypothesis that strictly cognitive perceptual adaptation does not occur, it is not clear to me why this shows that adaptation cannot involve concepts when they are tokened in perceptual contexts. In light of my earlier suggestion, we might grant, for example, that *Central Cognition* operates on principles that exclude adaptation phenomena while allowing that adaptation may be operative at the level of CSTM, or whatever similar mechanism serves to encode perceptual contents in respect of semantic content.

A second reason Block rejects the idea that the representation in question was conceptual comes from subjects' reports. For example, in relation to one experiment involving ambiguous figures (Peterson & Gibson, 1994) he claims that "the independence of what the subjects know suggests the adaptation is not cognitive". Without going into too many details, what Block seems to have in mind here is that subjects in this particular paradigm arrived at certain particular visual interpretations of ambiguous figures based on their orientation, despite having previously been told of other competing interpretations. As the original paper puts it, "[k]nowledge regarding which shape was depicted by the high denotative region was insufficient to produce shape recognition contributions to figure– ground organization in the absence of an orientation dependent match to a representation of an object in memory (e.g., in the inverted condition)." However, it is not clear why this should tell against the idea that subjects' performance in the experiment involved the application of concepts, especially in light of the kind of CSTM account of categorical perception sketched earlier. In particular, given that the mechanisms of conceptualization in perception presumably operate fairly rapidly and autonomously, the mere fact that subjects had prior knowledge of the different interpretations of the presented images may not suffice to ensure that they did *in fact* conceptualize the relevant shapes in respect of that knowledge on any given trial.

I would suggest, then, that while Block's arguments may tell against interpreting all cases of categorical perception in terms of Central Cognition, they do not allow us to adjudicate between the perceptual and CSTM-based accounts. Are there any other reasons why we might be skeptical that categorical perception could be carried out by perception alone? While I do not take the evidence to be decisive, I will now offer one brief passing consideration.

My challenge concerns the limitations of sensory representation. Elsewhere (Block,

2015) suggests that sensory representations possess a *pictorial* mode of representation.[20] This is not to suggest that mental icons are like internal photographs. A better comparison, one explicitly drawn by Block, may to be the representational forms open to painters; thus Block observes that "an impressionist painter might represent a hand in broad brush strokes that do not explicitly represent the number of fingers or whether one of them has a ring."

Even granting, however, that sensory representations can possess at a high degree of indeterminacy of content, it seems hard to see how their representational range can extend to all of the properties seemingly involved in categorical perception. In particular, it seems that the pictorial content of certain kinds of iconic representation drastically underdetermines their total representational content, and instead represent in ways that rely on previous acquaintance with appropriate stimuli and contexts in order for their meaning to be clear. Consider, for example, Fig. 3b. I take that this therefore constitutes at least a prima facie instance of perceiving someone *as* drunk.

[20] There are various ways to unpack this claim. One might suggest, for example, that pictorial representations constitutively involve analogue forms of representation (Dretske, 1981), represent in virtue of possessing isomorphisms to the representatum (Gombrich, 1983), or are subject to the "picture principle" according to which every part of a picture represents some part of whatever it is that is represented by the whole (Fodor, 2007). However, it is not clear which of these claims Block is committed to.

**Fig. 3b.** A cartoon of a drunk.

However, the distinctively *pictorial* features of the image– the man's red nose, his bulging eye, the glass in his hand – are also compatible with many other interpretations. Someone with a different history of perceptual learning might, for example, interpret the picture as showing a man with a bad cold drinking cough medicine, or a person who has just been in a fight who has managed to collect his spilled blood in a glass. Other cues to meaning in this picture are provided by the intrinsically meaningless 'emanata', the symbols above the figure's head that indicate his inebriation (see Fig. 3c).

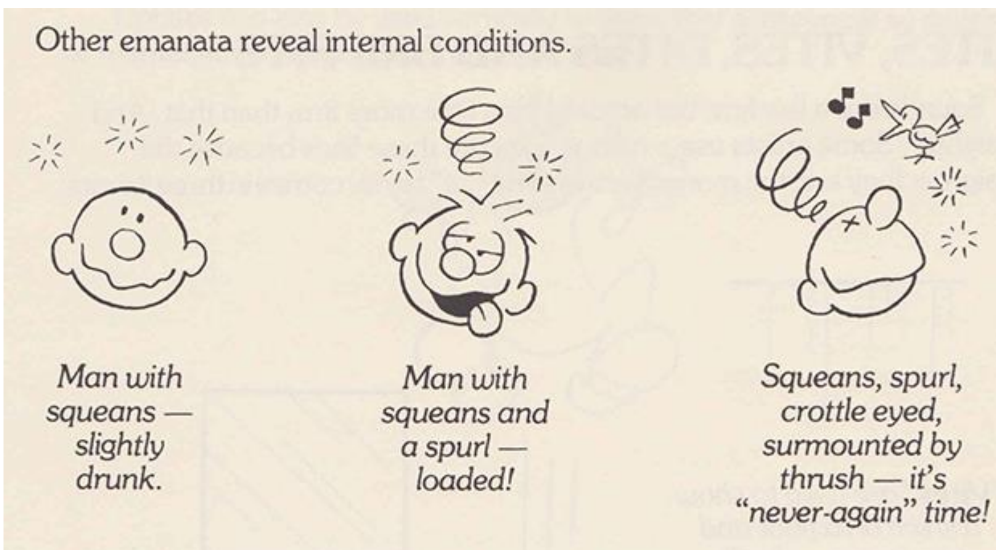



**Fig. 3c.** The role of 'emanata' in depicting drunkenness. From the Lexicon of Comicana (Walker, 1980).

I do not wish overintellectualize the process of understanding cartoons: we are certainly not aware of making any conscious inferences about the content of Fig. 3b, and we do not consciously *use* symbols like the emanata in Fig. 3c in a reflective manner (indeed, it is easy to fail to consciously notice them entirely). Nonetheless, thanks to most readers' familiarity with the relevant representational forms and conventions, these cues enable us to effortlessly represent the artists' desired content, in spite of the fact that this content is underdetermined by the purely *pictorial* aspects of the image in question. This suggests that, to the extent that properties like 'looking drunk' feature among the contents of categorical perception, they rely at least in part on some non-pictorial mode of representation.

Of course, Block may simply deny that there is perceptual-level representation of properties like 'being drunk'. However, it seems plausible that many of the same considerations will apply to lower-level contents such as 'predator' or 'shelter', which Block (2014a) and Burge (2010b) do suggest are among the contents of perception. The challenge here, then, is that it is hard to see how a creature could become aware of an entity as falling under one of these categories solely in virtue of the pictorial features of a given representation. We can imagine, for example, that two organisms may have visual representation of an animal (say, a pig) that have all their sensory features (color, shape, size, and so on) in common, but whereas one represents it as a predator, the other represents it as prey. Without appealing to non-sensory representational contents (such as inferential role; see, for example, Shea, 2011), it is hard to see how this difference in representational content can be accommodated.[21]

Admittedly, it is not clear whether Block and Burge take more sophisticated representational contents like 'predator' to possess a specifically pictorial mode of representation. However, if they nonetheless wish to insist that these forms of categorical perception are carried out by strictly sensory processes as opposed to conceptual (or otherwise semantic) mechanisms (as Block in particular seems keen to do), they must offer some further positive account that explains how strictly sensory modes of conscious experience can make a

---

It is of course open to theorists like Block and Burge to insist that the analogy I have drawn between sensory representation and literal pictures is misleading. However, if the analogy with pictures is inadequate, and subject aware of one kind of content rather than another in the kinds of cases discussed above.[22]

*(ii) Is categorical perception just Central Cognitive belief?*

Thus far, I have given some reasons for doubting that categorical perception can be adequately explained just in terms of perceptual processing. I now wish to consider the possibility that categorical perception might be a function of Central Cognition. The paradigm cases of Central Cognitive processing are conscious cognitive processes such as reflecting, judging, and deciding. Thus one seemingly promising way to cash out categorical perception in terms of Central Cognition would be to claim that it consists in the formation of rapid thoughts or judgments about the identity of an object in perception (see also Chapter 5).

I will now briefly review some reasons why I am also skeptical about this approach. Note in proceeding, however, that I wish to leave open the possibility that categorical perception – which I suggest involves encoding in CSTM – may indeed have a belief-like propositional structure (see 4.4). I also do not wish to foreclose the possibility that there could be (perhaps non-conscious) forms of belief or judgment that are realized outside of Central Cognition but are not anything to do with categorical perception. The specific thesis that I wish to examine, then, is the idea that categorical perception may consist in a conscious judgment or thought realized by Central Cognition.

With this in mind, I will note a few features of categorical perception that arguably

[22] It is worth stressing that the kind of worry about underdetermination I am raising here is distinct from the specific underdetermination problem discussed by Burge (2010, 2015) in relation to perception. In these contexts, Burge is concerned with the scientific problem of how visual systems can come to represent certain kinds of property in the environment. By contrast, my concern is with the more philosophical problem of how perceptual *experience* of a representation just in respect of its sensory features could result in our becoming aware of an object in respect of properties that are underdetermined by those features. Thus even if we have a scientific account of how Burge's perceptual constancies are achieved in the brain, we still require a philosophical account of how we are made aware of categorical properties in experience, and it is far from clear to me that this can be done just in terms of the sensory features of representation.

create some cleavage between perceiving-as and the typical instances of conscious judging and thinking. The first point is that categorical perception is relatively insensitive to correction by background beliefs: one can see an object as *F* even while one consciously believes or judges that the object is not-F. A simple illustration of this comes from cases of known illusion, notably discussed by Plato.

> The same magnitude, I presume, viewed from near and from far does not appear equal… And the same things appear bent and straight to those who view them in water and out, or concave and convex, owing to similar errors of vision about colors…And did we not say that it is impossible for the same thing at one time to hold contradictory opinions about the same thing? And we were right in affirming that. The part of the soul, then, that opines in contradiction of measurement could not be the same with that which conforms to it." (Republic, 602c-e; translated by Bloom, 1968)

Plato's examples relate just to shapes and colors, but there are examples of known illusion that involve misperceiving something in respect of higher-level categories, notably in the phenomenon of pareidolia (a kind of illusion in which we seem to perceive a pattern or object when none is present). Simple examples include seeing faces in clouds, trees, or spills. Of course, sometimes we say that something looks like something else without categorical perception truly being involved (as when I say that a friend has a passing resemblance to a famous actor). On other occasions, however, pareidolic illusions can be powerful and irresistible, as in the example given in Fig. 3d. What is especially important about such cases is that even when we know that we are the victims of an illusion, the categorical representation persists: no matter how firmly we tell ourselves that we are not looking at an eye in Fig. 3d, we cannot help but see it as one. Nor is our inability to adjust how we perceive the image in such cases idiosyncratic or self-interested: we *all* see the plughole as an eye, despite many differences in our background beliefs, and we do so in spite of the fact that no cherished beliefs about ourselves or the world would be threatened by not doing so.

**Fig. 3d.** an example of pareidolia (a plughole that looks like an eye)

The insensitivity of categorical perception to our beliefs provides us, I would suggest, with some initial reason to question the idea that categorical perception may just be a form of Central Cognitive processing, at least insofar as we take such processes to be typically amenable to correction by background information (such as those discussed in Fodor, 1983).

However, it would be to move too fast to claim that this was decisive evidence against the idea that categorical perception involved Central Cognitive beliefs. For one, there are many beliefs that do not seem sensitive to correction via background knowledge. Famous examples come from cases of cognitive dissonance, in which individuals willingly disregard information that threatens their ego or cherished beliefs. Nonetheless, these cases seem quite different from the kind of case shown in Fig. 3d. In this case, our inability to adjust how we perceive this image is not idiosyncratic or self-interested: we *all* see the plughole as an eye, despite many differences in our background beliefs, and we do so in spite of the fact that no cherished beliefs about ourselves or the world would be threatened by failing to do so.

Not all contexts in which failures of updating belief involve this kind of idiosyncratic or self-interested factors, however. There are many more mundane contexts in which people fail to properly update old beliefs in light of new information (Mandelbaum 2013, Egan 2008), and that our beliefs may be fragmented or 'compartmentalized' (Stalnaker, 1984; Lewis, 1982). We may therefore have good reason to ascribe contradictory beliefs to individuals. However, it is far from clear whether these sorts of cases establish the existence of contradictory beliefs that are simultaneously active at the same level of cognitive processing.[23] Indeed, some approaches to understanding contradictory belief make explicit reference to the idea that at least some cases of contradictory beliefs are explained by the two beliefs' being realized by distinct psychological mechanisms (Sloman, 2002). By contrast, if we wish to defend a central cognitive account of categorical perception, we must claim that an individual simultaneously represents two contradictory states of affairs ("that is an eye" and "that is not an eye") at the same level of cognitive processing. I know of no independent evidence to support this proposal. On these grounds, then, I would suggest that phenomenon of known illusion in cases of categorical perception provides some reason for thinking that perceiving-as does not involves Central Cognitive beliefs.

A second feature of categorical perception that arguably marks it apart from canonical cases of belief comes from the fact that it seems to involve a further sensory element. Compare the case in which I see James *as* James to the case in which I see James and simultaneously form a conscious thought that he has taken the day off work. I suggested

[23] See, however, Ripley, 2011, for one possible instance of simultaneous contradictory beliefs. This case is complicated, however, by the fact that the relevant beliefs concern vague predicates. By contrast, in Fig. 3d, it seems that I see the plughole as unequivocally an eye, while believing (unequivocally) that it is not.

earlier that the former but not the latter constitutes a plausible case of categorical perception. One reason for this, I now wish to suggest, is that in the case where I see some *x* as being F, there is a particular object or event in my *sensory* experience to which the property of being F is bound or attributed. Thus when I close my eyes, or when James steps momentarily out of my field of vision, it is no longer the case that I am seeing James as James. By contrast, if I see James and form a rapid automatic judgment to the effect that he has taken the day off work, this thought can persist in my awareness independently of my particular sensory state.

Of course, there is a perfectly reasonable response to this, which is to suggest that this simply marks a difference between two kinds of content rather than two kinds of cognitive process. Thus thoughts that were bound up with sensory content we might *label* perspectival and those that lacked such content non-perspectival, but this would not commit us to a fundamental division in kinds of mental attitude. This is a reasonable proposal that I will not attempt to refute here. However, note that one natural explanation of how beliefs might come to have quite different kinds of content (that in fact differ in format) would be to suggest that they involve two different kinds of psychological processes. This way of thinking about categorical perception would thus be a natural fit for the kind of CSTM account of perceiving-as outlined earlier

A final set of considerations that arguably tell against the idea that we should understand categorical perception in terms of Central Cognitive beliefs comes from cases in which it is arguably reasonable to think that an individual possesses categorical perception but lacks the capacity for states like believing or judging, and vice-versa. The best candidate for the former sort of case, I believe, comes from non-human animals. I consider this proposal in more detail in Chapter 5, but, in short, there is some evidence to think categorical perception may be widespread in nature, perhaps even occurring in invertebrates. By contrast, there is still considerable debate as to which non-human animals possess the kind of sophisticated cognitive architecture that may be involved in the possession of propositional attitudes (see, e.g., Bermudez 2003). While the empirical data here are far from decisive, *if* there are creatures that can engage in categorical perception but lack the ability to form thoughts or beliefs via some mechanism like Central Cognition in humans, then the former must not depend on the latter.

In the other direction, there is arguably evidence for massive failure of categorical perception in the absence of broadly preserved Central Cognitive abilities in the condition known as associative agnosia (see Appendix 5 for further discussion). In short, patients with associative agnosia have a drastically diminished ability to recognize objects. Some associative agnosics cannot recognize anything at all beyond very simple geometrical structures, while others suffer more specific deficits to recognition involving particular semantic categories, such as selective deficits to animate or inanimate objects (De Busscher, 1955). However, they do not seem to suffer any impairments to canonical Central Cognitive processes, and do not seem to encounter any difficulties in daily life aside from those linked to their specific inability to recognize objects (Gottfried, 2011: Ch.11). This in turn suggests there may be a dissociation between Central Cognition and categorical perception, at least the neural level, which in turn might lend weight to the idea that they are subserved by distinct psychological mechanisms.

## 3.4 – The CSTM model of categorical perception

While I do not take any of the above considerations to be dispositive, I believe they constitute good reason to take seriously the idea that categorical perception relies on a psychological process distinct from both Central Cognition and perception proper. CSTM, as suggested earlier, seems an excellent candidate mechanism in this regard, so I will now briefly spell out in a little more detail how this might help us understand the data.

I argued in the previous chapter that CSTM functions as an intermediary between perception and cognition, and serves to classify sensory information into semantic categories prior to the involvement of Central Cognition. These semantic categories, I would suggest, may include some simple (possibly innate) representations of categories such as ‘predator’, ‘food’, and ‘danger’, but can also (as suggested by Potter’s work) include high-level categories such as ‘wedding’, ‘Donald Trump’, or ‘drunk’. The subsequent deployment of these categories in perception explains how we are able, for example, to see the man in Fig.3b as inebriated without the need for any conscious inference, while the fact that CSTM is a distinct and relatively autonomous process from Central Cognition would account for the fact that categorical perception is relatively insensitive to correction by our conscious thoughts and beliefs.

How do these categories get applied? I am inclined to regard this question as a straightforwardly scientific one, to be answered by vision scientists. However, leading approaches appeal to processes such as template- or prototype-matching or featural analysis (see, e.g., Ullman 1996 and Corrigan 2007 for a more detailed discussion of these approaches). I also wish to remain agnostic on whether the processes involved in determining the semantic category of a given sensory input are best construed as associative or inferential. Again, I take this to be a largely scientific question, although one that is perhaps hindered by the lack of a widely accepted philosophical account of what makes a given transition between contents a strictly inferential one.

What is critical for my purposes, however, is the claim while the inputs to these processes may depend on purely sensory information, their outputs serve to bind sensory information into a semantic category that consequently bears not merely sensory but *semantic* relations to other items in an individual's perceptual field (as suggested by Belke et al., 2008) as well as to their own goals and motivations, allowing them, for example, to recognize that an given object is relevant to their current task.

It is worth spelling out in a little more detail what this binding might amount to. The basic proposal I have in mind is that representations in CSTM may be a kind of hybrid representation, in which a conceptual 'label' is indexed to an underlying iconic representation. This may seem an obscure notion, but I would suggest that we are all familiar with these kinds of hybrid representational formats in the form of labelled maps and diagrams (Camp, 2007). Consider, for example, a seating chart, which lists the names of various people, but also indicates their position relative to one another. This seems to me to be a promising approach, accommodating the varied ways in which we can make use of perceptual experiences. Crucially, note that if this proposal about hybrid content is along the right lines, it might explain the observation in 3.4 that categorical perception always involves a sensory as well as a categorical component.

While the idea that information in CSTM has both semantic and sensory content is just one suggestion for how information might be encoded, I take it to be supported by the finding of Potter et al. (2014) that subjects' ability to recognize images in forced choice tasks depended on their first having detected that the image was presented under a conceptual label. While recognition would presumably depend on subjects' recall of the pictorial features of the image, detection under a conceptual label would depend on the category by which it was encoded. The fact that the two abilities were bound up with one another lends some weight to the idea that encoding in CSTM involves a hybrid format.

Moreover, I take it that this proposal has some empirical support. For example, experimental work by Grill-Spector and Kanwisher (2005) examined subjects' ability to detect, categorize, and identify visual stimuli. All tasks involved presenting items for varied brief intervals (ranging from 17-200ms). The detection task required subjects to decide whether an object (as opposed to a texture) had been presented. The categorization task used a similar methodology but required subjects to press a button to indicate whether an item from a target category (e.g., car vs. not-car). Finally, the identification task required subjects to indicate whether the item belonged to a within-category class (e.g., German Shepherd vs. some other dog). Grill-Spector and Kanwisher discovered that while subjects were slower on the fine-grained identification task, their subjects' accuracy and speed was just as fast for the *categorization* task as it was for the *detection* task. Summarizing their results, they note that "detection and categorization performance require the same amount of information and processing time" and that "[b]y the time subjects knew an image contained an object at all, they already knew its category."

This result supports the intuitive idea that the basic semantic information is, at least in many cases, present in perceptual experience from the first moment that a stimulus reaches awareness. Note also that subjects' relative slowness in fine-grained identification tasks as compared to basic categorization tasks also suggest that the relevant kind of categorization involves the application of relative *generic* semantic categories ("dog", "guitar", "car", and so on). Again, this is readily accommodated by the kind of account I have described.

Finally, note that by adopting the claim that categorical perception relies on a distinctive psychological mechanism, we can also account for cases such as associative agnosia, in which Central Cognition and perception proper seem to be left broadly intact even while categorical perception has been drastically impaired. Moreover, it seems possible that some non-human animals may possess a faculty that is broadly equivalent to CSTM even in the absence of mental architecture corresponding to Central Cognition. This may allow us to account for apparent cases of categorical perception in the absence of more robust cognitive states. This is a proposal I will return to in Chapter 5.

Before moving on, it is worth mentioning a few questions regarding which I wish to remain broadly agnostic for the time being. First, note that I have made no definite claims concerning the syntactic structure of conceptual information in CSTM. In particular, I have not given a definite answer to whether they possess full-blown propositional content ("that is a predator") as opposed to some simpler structure, such as a concept bound to a referring element ("that-predator"). Though I lean toward the latter, I see no immediate way of settling the question.

Second, I have not yet said anything substantive concerning the relation between categorical perception and the mechanisms of conscious experience. This will be the subject matter of the next chapter. Relatedly, although I suggested that the typical function of CSTM may involve making states available to Central Cognition, I do not wish to foreclose the possibility that some states in CSTM may be cognitively inaccessible. Again, this is a suggestion that will be picked up in the next chapter.

Finally, note I have made only modest commitments regarding the variety of contents that can be applied in CSTM. I do not foreclose the possibility that *any* conceptual content (including highly abstract contents such as "being a recession" or "being communist") could potentially feature among the contents of CSTM. However, I also think there might be good reasons (perhaps reflecting the mechanisms of perceptual learning) for thinking that the contents of CSTM might be constrained to representing the features of more concrete objects and properties.

## 3.5 – The phenomenology of categorical perception

Thus far I have given an account of CSTM that claims it is a distinct psychological mechanism, and one which may enable us to explain categorical perception. In the last two parts of this chapter, I wish to outline how this framework might be usefully applied to two other related debates, one concerning high-level phenomenology, and the other concerning cognitive penetration.

The high-level phenomenology debate is concerned with the claim that phenomenology non-derivatively represents high-level properties, a position also known as the Rich Content View (Bayne 2009, Siegel, 2010). This proposal holds that there are phenomenal properties associated not just with low-level features such as color and shape but a range of high-level content such as natural kinds, individuals, and causal relations. Thus it has been suggested that there may be distinctive kinds of phenomenal property associated with the perception of particular classes of objects: *looking like a bicycle*, for example, or *looking like a pine tree*.

The Rich Content View is highly controversial. Some theorists, for example, take a *deflationary* approach to high-level phenomenal content, arguing that the differences in such cases consist in differences specifiable in terms of the modulation of low-level contents (see, for example, Carruthers & Veillet, 2011). This kind of account can be spelled out in terms of differential allocations of attention before and after perceptual learning (with resulting shifts in phenomenal character; see, e.g., Fuller, Ling, & Carrasco, 2004) or in terms of shifts in the way that my perceptions are structured. Thus prior to learning how to identify Cyrillic letters, I might see them as a jumbled mess, and, without careful attention, struggle to pick out individual characters. Once I have learned to read them, however, I can more easily structure my experiences into individual items. Similarly, in the case of learning to identify an albatross, I might more rapidly attend to certain visual features than others, such as its distinctively long bill.

While such strategies may explain many apparent cases of high-level phenomenology, I am nonetheless sympathetic to Siegel's broader claim that high-level properties in perceptual experience may contribute to the phenomenal character of the experience independent of the phenomenal character of our low-level perceptual properties. While I will not attempt to defend these sympathies in detail here, one example may serve as an example of why I take the proposal to be appealing and the strict deflationary approach to be inadequate. Consider the Thatcher Illusion (Fig. 3e). In this effect, what we in fact see is an inverted face with a vertically rotated mouth, though it strikes us, at least to begin with, as a normal smiling face. However, gazing at the image a little longer, it quickly becomes clear that there is something wrong with the image; in other words, we see it as exhibiting some distortion (even if we cannot immediately pin down what it is). However, it is doubtful whether the strictly *perceptual* content of the image changes at all: it is not as though we see the smile 'flip' orientation. Instead, throughout the time when are looking at the image, our representation of the low level features of image is presumably veridical; that is, we see the mouth as being at the same orientation (inverted) throughout. What changes is some other aspect of our overall experience, specifically whether we see the fact as normal or distorted. This suggests a phenomenological role for categorical perception independent of how we represent the low-level features of the image.

**Fig. 3e.** The Thatcher Illusion (Thompson, 1980). This image initially looks normal, but staring at it, it quickly starts to look disturbing, even though low level features remain the same.

Here, then, is an initial place where CSTM view might be put to good work. On the assumption that categorical perception involves semantic encoding in CSTM, it might be the case that this encoding, when conscious, made a distinctive contribution to phenomenology. Applied to the Thatcher illusion above, for example, the phenomenological shift in my experience may be caused by a change in the way that I conceptualize the image in CSTM; very simply, for example, I might go from conceptualizing the face as "normal" to "strange", with the result that my perceptual experience instantiates different high-level phenomenal properties.

Independently of these specific suggestions, I would draw attention to the fact that there is, to my knowledge, no other candidate mechanism of short-term memory besides CSTM that might otherwise set Siegel's account on a firm empirical setting. In particular, note that Siegel strenuously denies that rich content is a matter of the kind of high-level cognitive phenomenology that is sometimes (e.g., Siewert 2011) claimed to accompany states such as reflection and attentive thought. For example, in the following passage, she contrasts the (strictly cognitive) phenomenology of carefully reading a sentence with merely glimpsing text in passing.

> Contrast this phenomenology [of reading] with that of being bombarded by pictures and captions on billboards along the highway. This seems a visual analog of the blare of a loud television, or a fellow passenger's inane cell-phone conversation: understanding the text on the billboard as you drive by isn't a deliberate affair; rather… it just happens… The advertisers would doubtless be happy if you lingered over every billboard's message, but no such event need occur in order for you to take in the semantic properties of the text as you whiz by. This suggests that the taking in can be merely sensory. (Siegel, 2010: 108)

Siegel is drawing a contrast here between our fleeting *perceptual* experience of semantic content that happens effortlessly and automatically and the more reflective *cognitive* awareness of such content that comes when we notice or pay close attention to things. Note that this latter presumably involves Central Cognition, thus if her contrast is to be plausible we must appeal to some other mechanism. However, for reasons given earlier, perception proper seems highly limited in the kind of properties it can represent.

Siegel may wish to resist the idea that CSTM underpins high-level phenomenology on the grounds that it involves the application of concepts and should consequently be considered a broadly *cognitive* process. However, insofar as the CSTM account provided takes categorical perception to occur independently of central cognitive processes, it can also accommodate the kind of distinction that she wishes to maintain here between the phenomenology of perceptual experience and the phenomenology of reflective thought (or other paradigmatically cognitive activities). Indeed, the claim that conscious encoding in CSTM contributes to the phenomenology of perceptual experience is compatible with both the acceptance *and* the denial of the contentious claim that there is also a further kind of phenomenology associated with non-perceptually directed conscious thought.

I would suggest, then, that the CSTM account of categorical perception offers a promising route for Siegel and other defenders of high-level phenomenology. However, the CSTM theorist themselves is certainly not tied to this strategy. Indeed, the account as given in the previous sections is, I take it, broadly compatible with a range of different approaches to high-level phenomenology, including radically deflationary ones.

Before moving on, however, I wish to mention one further account of high-level phenomenology that I believe may offer the CSTM theorist an attractive alternative to Siegel's view while allowing that the manner in which sensory information is conceputally encoded in CSTM makes a distinctive contribution to the phenomenal character of experience. This is the imagistic account of high-level phenomenology given by Prinz (2011a). While Prinz's view can broadly be considered a deflationary approach insofar as it denies a proprietary phenomenology of high-level content, it importantly differs from the approach taken by other deflationary theorists (such as Carruthers & Veillet, 2011) insofar as it explains the phenomenological contrasts found in apparent cases of high level phenomenology not in terms of attention to low-level features but rather the activation of *mental imagery*. Thus, considering the kind of phenomenological shift involved in bistable imagery such as the duck-rabbit, Prinz states as follows.

> When we interpret an image as of a duck, we gain access to duck knowledge stored unconsciously in long-term memory, and some of this knowledge may bubble forth in the

> form of mental imagery. It is possible, for example, that when construing the duck-rabbit as a duck, viewers faintly imagine a duck's body and other features. (Prinz, 2011a, 183).

Prinz goes on to cite evidence that monochrome images of everyday objects are sometimes perceived as colored (Hansen et al., 2006) and that visually presented animal names prime animal sounds, and vice versa (Orgs et al., 2006).

A further line of support for Prinz's claims might be found in the literature on 'boundary extension' (Intraub & Richardson, 1989).[24] In short, this is a phenomenon observed when subjects are shown images and asked to redraw them afterwards after variable delays. A consistent finding reported in more than 90% of subjects is that participants' drawings *extend* the field of view of images and tend to feature complete renditions of the objects that were partially occluded when originally presented (see Fig. 3f). There are a variety of interpretations here. For example, one possibility is that subjects are merely forgetting details of what they had seen and distorting their memory of the original picture. However, contrary to this interpretation, the effect seems to get *weaker* rather than stronger in time. This suggests that it may in fact be a perceptual effect, reflecting, perhaps, the activation of appropriate mental imagery during original presentation. This again might lend further weight to Prinz's approach to high-level phenomenology.

[24] Thanks to Nicholas Porot for bringing this literature to my attention.

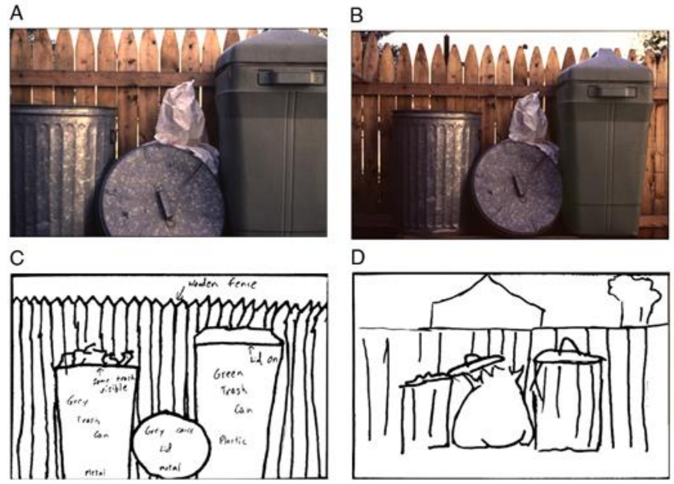



**Fig. 3f.** Subjects presented with the images on the top row later drew the pictures below when asked to create a picture of what they had seen (Intraub & Richardson, 1989)

I will not attempt an extended defense of Prinz's account here. I only wish to note that this kind of approach may again offer a natural fit for the CSTM view. It may be the case, for example, that when sensory information is encoded in respect of a concept in CSTM, this causes the activation of category-appropriate imagery, such as stored images of ducks. Note, moreover, that one the studies cited by Prinz (Orgs et al., 2006) appeals to category-appropriate priming of *across* different sensory modalities, suggesting that the relevant mechanism of priming may have been conceptually mediated.

To summarize, then, the CSTM account of categorical perception given in this chapter offers a number of promising strategies for account for putative cases of high-level phenomenology. Moreover, for theorists such as Siegel and Bayne who wish to maintain that perceptual experience involves high-level phenomenal properties that are nonetheless importantly different from the phenomenology of Central Cognition proper, appeal to CSTM or some similar mechanism may be indispensable if they wish to spell out their account in terms of empirically-supported psychological processes.

## 3.6 – CSTM and the cognitive penetration debate

One final question about the perception-cognition boundary where my account may be of relevance concerns cognitive penetration. Very roughly, the debate concerns the extent to which cognitive processes can affect (or 'penetrate') perceptual processing, and is closely related to the observations made earlier regarding the encapsulation of perception from central cognitive processes. The debate is potentially of great relevance to a number of questions not merely in psychology and cognitive science, but also the epistemology of perception.

To give a simple example of the kind of case at issue, imagine two spectators are watching a tennis match between Djokovic and Murray, and Murray's serve is ruled to have clipped the net. Spectator A, a Djokovic supporter, agrees with the Umpire's decision, and insists they saw the ball clip the net, but Spectator B, a Murray supporter, claims that he clearly saw the ball skim over the net without touching it. Assume further than both supporters have good vision and had a clear line of sight to the net.

We can now ask whether their respective partialities to the two players led them to see different events unfold. If the mechanisms of visual awareness are encapsulated from subjects' attitudes and desires, then the answer to this will be no, and their different judgments in the two cases are likely due to a bias in judgment on the part of one or both of them, for which they can be held accountable as epistemic agents. On the other hand, if their attitudes towards the two players did influence their visual experience prior to the operation of cognitive processes, then both players could be responding fully rationally given the evidence at their disposal, making this perhaps a case of blameless disagreement. This example is a little simplistic, but should give the reader an idea of the kind of case at the heart of the debate.

A copious amount of experimental data (dating back to early work by psychologists such as Perky, 1910, but proliferating especially in the last two decades) has been produced in support of the idea that there are indeed 'top-down' effects on perception, which in turn has been taken to suggest that our thoughts, beliefs, and desires can significantly how the world appears to us. Some of this data has been taken to suggest that perceptual learning can influence the way we perceive even low-level properties such as hues (Goldstone, 1995; Hansen et al., 2006) and the perceived size of objects (Bruner, 1957). For example, one study by Levin and Banaji (2006) suggested that faces categorized as racially black rather than white are perceived as having a darker skin tone, even when luminance is controlled for, while a study by Den Daas et al. (2013) suggested that men primed with sexual thoughts judge women's breasts to be larger than men who had not been thus primed.

A further important line of research concerns apparent distance and the influence of action on perception. Thus one influential study (Profitt et al., 2003) suggested that subjects wearing a heavy backpack gauged distances to be longer than subjects who were unencumbered, while an experiment by Witt et al. (2004) showed that subjects instructed to throw a heavy rather than light ball made greater estimates of the distance to a target, leading the authors of the study to suggest that the expectation of increased exertion involved in throwing the ball may actually affected how far away the target looked.

This is a very small sample of the rich array of data that has been claimed in support of the existence of top-down effects on perception. However, such conclusions are increasingly being contested on methodological and theoretical grounds. Thus Firestone and Scholl (2015) single out a range of factors that they take to count against interpreting such data in terms of genuine top-down effects, including failures of experiments to account for the 'El Greco fallacy' (Firestone, 2013), the role of memory in speeding perceptual processing, 'demand effects' in which subjects' reports are influenced by their beliefs about the purpose of the experiment, and attentional effects (Firestone & Scholl, 2015).

This is a complex and rich debate, and I will not attempt to offer any broad conclusions about it here. However, I do wish briefly to consider how my model of categorical perception as involving a distinct mechanism in the form of CSTM may be of relevance to the dispute. In the broadest terms, the hypothesis that CSTM underwrites categorical perception and operates as an intermediary between perception and Central Cognition suggests there may be compromise positions available in the interpretation of alleged cases of top-down effects on perception. For example, Firestone and Scholl claim that many such results fail to properly distinguish *perception* from *judgment*, claiming that, in many cases, experimentalists' results can be interpreted purely in terms of strictly cognitive effects rather than as involving effects on perceptual experience. In this spirit, they claim that it is important to distinguish perception from judgment carefully when assessing whether top-down effects are present, and to avoid blurring the lines between the two. As they warn us, "many papers in this literature advert to effects on "perceptual judgment"… which can only invite confusion about this foundational distinction."

However, if CSTM is a distinct psychological process operative between perception and Central Cognition, then appeals to processes such as 'perceptual judgments' may be better founded than Firestone and Scholl seem to think. While I have suggested that CSTM may be relatively autonomous (insofar as, for example, forms of categorical perception persist even when we know them to be illusory), it seems an open empirical possibility that our beliefs, desires, and other cognitive attitudes can have more subtle influences on how and when percepts are encoded in respect of specific semantic categories, even if there is no direct cognitive influence on perception proper (thus also respecting the claim that early vision is encapsulated; see Pylyshyn, 1999). This might in turn suggest that certain alleged top-down effects are neither strictly perceptual nor strictly cognitive, insofar as they ensue neither from low-level perceptual processes nor from a subject's central cognitive processes.

If this were the case, then it would mean that while the strictly sensory elements of perception (such as basic shades and contours) were not directly penetrated by cognition, the conceptual contents of perception (such as high level color and shape *concepts*) might indeed by thus penetrated. Thus it is possible, for example, that subjects in the study of Witt el. al. who were holding a heavy ball encoded their perceptual experiences in CSTM under different semantic categories than those who were holding lighter balls, thereby (perhaps) leading them to judge that the targets were further away.

Would this count as genuine cognitive penetration? This depends on one's broader theoretical commitments and what one takes to be at stake in the debate to begin with. Certainly, as noted, this possibility is compatible with a model of the human mind which allowed for a high-degree of encapsulation of core perceptual systems. On the other hand, insofar as CSTM seems to operate automatically and involuntarily, it might allow for genuine blameless disagreement in cases where subjects claimed to have experienced the same event differently. Perhaps our two spectators in the tennis case described earlier really had experiences with different contents, albeit with the differences applying at the conceptual rather than strictly perceptual level of processing.

One might also press the question of whether such possible cases would involve phenomenologically different experiences. In other words, could cognitive penetration of CSTM (but not perception) lead two subjects to have experiences that really *looked* or *sounded* different? The answer to this question will depend on the broader issues about how CSTM relates to phenomenal character outlined in the previous section. If, for example, we take CSTM to involve high-level phenomenology (like that suggested Siegel and Bayne) we might be inclined to answer affirmatively. Even if we adopt an imagistic account (like that of Prinz, 2011a) of the phenomenology of CSTM, we may also have grounds for thinking that perceptual experience itself could be affected by the manner of encoding in CSTM (cf. MacPherson, 2012).

The account suggested above is highly speculative: further empirical and theoretical work is required to establish, for example, the extent to which rapid and automatic processes of perceptual categorization (as exemplified by Potter's work) can be indeed influenced by subjects' broader beliefs and desires. However, I take it that this constitutes a debate in which CSTM may serve to create new theoretical possibilities and suggest further research.

## 3.7 – Conclusion: CSTM as apperception?

In this chapter I have suggested that by expanding our model of the mind to accommodate CSTM, we can more readily explain the puzzling phenomenon of categorical perception, with potential consequences for other debates such as those concerning high-level phenomenology and cognitive penetration. If these proposals are correct, then philosophy of mind and cognitive may have been missing out something important for all these years.

The very audaciousness of this proposal may seem to count against it. Can we really suppose that generations of philosophers and scientists would overlook a fundamental part of the mind? In fact, I think this worry is less acute than it first appears: there are many reasons why CSTM may have been overlooked. For example, there is still a proliferation of competing models in all the fields in which CSTM is of relevance, not merely in philosophy of mind but also in the psychology of memory. This makes it harder for new forms of memory to become generally accepted. Moreover, it is extremely experimentally challenging to disassociate short-term mechanisms of conceptual memory from forms of sensory memory and from working memory, and it is only in Potter's most recent work that the clearest evidence of CSTM as a distinct faculty has emerged.

Additionally, however, note that the kind of process that I take to be constituted by CSTM is not strictly new. Indeed, numerous thinkers and philosophers have previously insisted upon a mechanism intermediate between perception and understanding in the form of what they called *apperception*. This term was famously used by both Leibniz and Kant in their otherwise very different accounts of the mind to refer to processes by which the deliverances of the senses were made available to conscious understanding. More recently, the term was used by 19$^{th}$ century psychologists such as Wundt, Herbart, and James to refer a process distinct from both perception and understanding that was operative in perceptual experience and played a key role in making perception intelligible. Thus James (1900), with intriguing parallels to the kind of theory I have laid out in this chapter, claims that "we never get an experience that remains for us completely nondescript: it always reminds of something similar in quality, or of some context that might have surrounded it before, and which it now in some way suggests… We conceive the impression in some definite way. … This way of taking in the object is the process of apperception."

There are many differences and subtleties among these thinkers' accounts of apperception which I will not explore here: this is not the place for historical exegesis, nor do I wish to claim that the kind of account I have presented would necessarily be endorsed by these thinkers. However, I do take it to be of interest that previous generations of philosophers and psychologists believed that a further faculty was required to bridge perception and conscious understanding.

I hope that the basic framework of the CSTM model of perception is now in place. However, one set of important questions that I have assiduously dodged in this chapter concerns the relationship between CSTM and consciousness. Is CSTM always conscious? Could CSTM even contribute to a deeper understanding of the mechanisms that enable conscious experience in the first place? These are the questions I will now turn to in the coming chapter.

# CHAPTER 4: CSTM and Theories of Consciousness



## 4.1 – Introduction

In the previous chapter, I argued that CSTM may be the mechanism that underlies categorical perception. Additionally, I suggested that if this proposal is on the right lines, then the way information gets encoded in CSTM may make a difference to conscious experience, either directly through a proprietary phenomenology of its own, or indirectly through the activation of stored imagery and associations.

I now wish consider whether CSTM may be of relevance to another debate related to consciousness, namely that concerning *theories* of consciousness. This area of research, as noted in Chapter 1, is concerned with identifying the psychological processes that result in information in the brain becoming conscious. Most theorists accept that there are various forms of unconscious perception, and even philosophers who question this (such as Phillips,

2015) allow that there are forms of information processing in the brain that do not directly contribute to conscious experience.

Explaining why some states are conscious and others are not thus provides the challenge that a theory of consciousness aims to address. In this chapter, I will explore how CSTM might contribute to this debate. I begin in 4.2 by spelling out in a little more detail the tripartite schema for thinking about theories of consciousness that I first presented in Chapter 1. I then go on to in 4.3 to suggest some of main desiderata for a theory of consciousness, noting that none of the major theories can satisfy all of these. In 4.4, I sketch a theory of conscious experience with CSTM at its core called the Workspace-Plus model, and explore various ways this might be cashed out. In 4.5, I examine the advantages of this sort of view, and argue that it can enable us to satisfy all of these desiderata discussed earlier. I go on in 4.6 to consider some of problems with the Workspace-Plus approach. However, detailed discussion of one of these problems, namely how to make sense of animal consciousness in relation to the Workspace-Plus account, will be largely deferred until Chapter 5. Finally, in 4.7 I consider a recent proposal by Peter Carruthers for how we should rethink the notion of cognitive access, and consider its viability and its relation to my own account.

Before proceeding, it is worth briefly noting that arguments of this chapter concerning the *mechanisms* of conscious experience are at least to some extent distinct from the issues considered in the previous chapter concerning the *contents* of conscious experience. For example, one might allow that the contents of CSTM make a distinctive contribution to conscious experience while insisting that they do so only in virtue of some further general mechanism of conscious. This could be, for example, attention (Prinz, 2012), availability for the formation of higher-order thought (Carruthers, 2005) or actually being the target of such a thought (cRosenthal, 2005). Thus even if one rejects one of the key claims of this chapter, namely that CSTM can be conscious in its own right, one might still endorse the claim that the contents of CSTM can contribute to conscious experience by virtue of some further mechanism.

## 4.2 – Spelling out the major theories of consciousness

In Chapter 1, I suggested that many of the major theories of consciousness can be broken down into a tripartite scheme that distinguishes them according to the role they take cognitive access to play in generating conscious experience. *Access* theories as I define them take cognitive access (that is, actual access by Central Cognition in some form or another) to be constitutive of conscious experience: any contents that are not cognitively accessed by a subject at a time are ipso facto non-conscious. Access-independent theories, by contrast, deny that any such link must obtain; even if accessibility to Central Cognition may make it more likely that a given sensory state becomes conscious, no such relation is essential. As Block (2014c) puts it, "consciousness greases the wheels of cognitive access but [consciousness] can obtain without it." Finally, accessibility theories strike a middle ground, holding that a state must be available to the mechanisms of cognitive access while allowing that states can be conscious while remaining contingently non-accessed. I will now briefly spell out some of the major approaches within this framework, illustrating their advantages and limitations.

*(i) Access theories*

Consider access theories first. A widespread (perhaps dominant) view among cognitive scientists is that consciousness can be understood as a kind of information-sharing in the brain, where this is mediated by some form of central cognitive process. This model of consciousness as information-sharing was popularized by Baars' (1988) global workspace theory, and is explicitly or implicitly assumed by many authors. As noted in Chapter 1, this this includes both Alan Baddeley (2003: 836) and Nelson Cowan (2001: 91), who suggest that consciousness arises in Central Cognition, and specifically via encoding in working memory.[25] A similar view has also long been defended by Dennett, who claimed that we should consider a state conscious to the extent that it has wide-reaching effects on cognitive processing; as he memorably puts it, "consciousness is cerebral celebrity" (Dennett, 1993).

Particularly influential among contemporary access theories is the theory advanced by Stan Dehaene (2014) and Lionel Naccache (Dehaene & Naccache, 2001). These authors offer a detailed picture of how consciousness arises in the brain, specifically taking it to occur when information is broadcast within a system termed The Global Neuronal Workspace. Using a wide array of empirical data, they offer powerful arguments in support of the idea that information only becomes conscious when it triggers a 'global ignition', causing synchronization of firings across different brain regions linked via distinctive pyramidal neurons.

Crucially, they take the function of this workspace to consist in making information immediately available for executive tasks such as report and voluntary action, and exclude the possibility of consciousness in cases where subjects claim to be unaware of stimuli. Based on considerations about subjects' limited ability to give report on multiple stimuli at a time, as well as architectural limitations in the global neuronal workspace, they also take the contents of conscious experience to be highly limited. Thus Dehaene and Naccache state that "studies indicate that when attention is distributed, temporal and prefrontal regions act as

[25] As noted in Chapter 1, I will not discuss evidence for a distinctive form of unconscious working memory in the main body of this dissertation. Instead, see Appendix 3.

capacity-limited filters and cannot represent more than a few objects" (2001).

In reading the mass of evidence Dehaene has accumulated for his position, it is tempting to think that a solution to the problem of consciousness is within reach. However, we should remain open to the possibility that the kind of neural psychological processes Dehaene has identified reflect some *subtype* of consciousness rather than consciousness tout court, perhaps indicating just the presence of cognitive access (Godfrey-Smith, 2016).

### *(ii) Access-independent theories*

I wish now to turn to a closer examination of what I term access-independent theories. These theories all claim that conscious perception does not constitutively depend on the limited mechanisms of cognitive access. As noted, Block is the most prominent defender of this sort of view, but his position is endorsed by Burge (2007) and a similar view is advanced by neuroscientist Victor Lamme (2010). According to these theorists, we should not assume that consciousness involves report or cognitive access, and we should instead treat it as a biological or psychological property that may or may not be reliably coinstantiated with properties like cognitive access and availability for report (Shea, 2012).

One of the important arguments for access-independent theories is phenomenological, appealing to the richness of experience (as discussed in more detail below). Another important and related strand of evidence in support of the idea that consciousness can be dissociated from cognitive access comes from empirical work on the partial report paradigms mentioned in Chapter 1. For example, Block notes that in several experiments demonstrating the high capacity of sensory memory (Sperling 1960, Landman 2003), subjects are under the impression of having seen more items than they were able to report. Thus, he claims that if we are ‘taking what subjects say at face value’, we should conclude that ‘[w]hat is phenomenal but in a sense not accessible, is all the specific shapes of the rectangles’ (Block, 2007, referring specifically to the experiments of Landman et al., 2003).

Somewhat roughly, Block’s argument is as follows. It has been empirically demonstrated (in Sperling, 1960, and Landman et al., 2003, for example) that subjects can retain information in sensory memory about all the specific items in briefly presented arrays, and in such cases they themselves have the impression of having been conscious of all that information in its specific detail, even though they cannot report everything. On the most straightforward interpretation, then, we should take subjects’ at their word and adopt the view that at least some forms of sensory representation can be conscious even when not cognitively accessed. Specifically, Block suggests that representations in sensory memory may be the constitutive basis for visual experience, and are capable of giving rise to consciousness even when cognitive access cannot occur. A similar position is advanced by Victor Lamme (2010), who claims that the basis for conscious visual experience is localized recurrent loops of firing in sensory areas of the brain, which again might in principle occur in the absence of any possibility of cognitive access.[26]

*(iii) Accessibility theories*

[26] Block and Lamme’s interpretations of partial report paradigms have faced a variety of sophisticated and detailed replies. A common strategy among such responses has involved attempting to explain away the seemingly large capacity of phenomenal consciousness in terms of ‘generic’ or ‘fragmentary’ phenomenology, arguing that our seemingly rich visual world is compatible with conscious experience being based in strictly capacity-limited mechanisms like working memory (Grush, 2007; Kouider, Gardelle, Sackur, & Dupoux, 2010; Stazicker, 2011). Another original approach is offered by Phillips (2011) who suggests that we can understand partial report paradigms without overflow by rejecting the picture, typically assumed in the debate, of how experiences unfold over time.

Access-independent theories remain highly controversial, and bring with them some seemingly radical consequences. In particular, several philosophers have questioned Block and Lamme's accounts of consciousness on the grounds that they implausibly broaden the range of states that could potentially be conscious. A third kind of approach, accessibility theories, attempt to steer a course between cognitively unconstrained view of phenomenology provided by access-independent accounts and the grounded but phenomenologically austere approach of access-theory. Like access-independent theories, accessibility theories typically take the basis of perceptual experience to be sensory states. However, they part ways with access-independent theories in making the further claim that it is only when such states *become available* for cognitive access that they are conscious, whether or not they are in fact thus accessed.

Two prominent theories of this type are Carruthers' dispositionalist higher-order thought theory (2005) and Prinz's AIR theory (2012). Both theories claim it as a strength that they are especially well-placed to capture major intuitive aspects of our folk notion of consciousness – in particular, our sense that consciousness is phenomenologically richer than cognition, and that our conscious experiences are ones that we can reflect upon and form beliefs about. As Carruthers (2011) puts it, for example, his dispositionalist account allows us to "retain our belief in the *rich and integrated* nature of phenomenally conscious experience" (italics added).

The theories differ somewhat, however, on the precise mechanisms involved. On Carruthers' view, what makes a state conscious is its being available to mechanisms of higher-order thought, where this involves a state's being able to be globally broadcast in the sense defended by Baars (2002). Prinz's account, by contrast, claims instead that conscious states are those that are available to working memory, where a state's being available is a matter of its being *attended*. However, the notion of attention Prinz has in mind here arguably differ somewhat from the highly capacity-limited mechanism that, for example, Cowan (2001) has in mind. Rather, attention is taken to be a more functionally definable process, namely the mechanism or set of mechanisms that makes information available to working memory (Prinz 2012: 97), though one that Prinz claims has a specific neural realization in the form of Gamma Vectorwaves (Prinz, 2012: Ch. 4).

## 4.3 – Desiderata for a theory of consciousness

Thus far, I have provided a framework for categorizing many of the major theories of consciousness, although I should stress that this is not meant to be exhaustive of all approaches (see Appendix 4). I will now develop a case for the claim that none of these accounts can give us everything we want from a theory of consciousness. Instead, I will argue for a new approach that I term the *Workspace-Plus model*. In very brief summary, this theory extends the kind of Global Workspace account offered by Dehaene to include the information that is encoded in CSTM. On this view, then, cognitive access may be one form of consciousness, but it is not the only one: the conceptualized percepts briefly stored in CSTM are also conscious, and give rise to our rich but fleeting perceptual awareness of the world.

I will shortly spell out this view in more detail. However, in order to properly motivate it, I now wish to lay out a set of desiderata that a theory of consciousness should aim to fulfill. In addition to being a useful exercise in its own right, this will lay the ground for the account I will shortly offer by demonstrating some of the limitations of existing theories. By contrast, the account I will offer in 4.4 might be able to give us everything we are looking for.

*(i) Phenomenal adequacy*

An initial desideratum for a theory of consciousness is that it should be *phenomenologically adequate*. That is a vague demand, but the key idea is that a theory should be able to be squared, one way or another, with introspective evidence about the character or content of conscious experience.

This covers a wide range of phenomena, but the single most important phenomenological consideration in the theories of consciousness debate has almost certainly been the challenge posed by the apparent *richness* of subjective experience. Somewhat crudely, this is the assertion that our experience contains many elements and objects that we do not notice or consciously think about. For example, just imagine being in Times Square, surrounded by noise and lights and people and smells, and reflect on how little of your experience it seems like you actually cognitively process in any detail. Intuitively, it's tempting to think that there is a significant gap between the many things we experience and the relatively few things we actually notice or consciously think about. According to access-independent accounts like those of Lamme and Block, our intuition here is exactly right: our perceptual experience is much richer than (or *overflows*) any mechanisms of conscious cognition.

By contrast, the strict capacity limits of most central cognitive mechanisms like focal attention and working memory mean that richness is more or less incompatible with access theories of consciousness. We can focally attend to very few items at any given time, and high-level cognitive faculties like the global workspace are very limited in capacity. Dehaene states, for example, that “[t]he global workspace model claims that, at any given moment, out of this enormous potential set, a single object of thought gets selected and becomes the focus of our consciousness”, and that “[w]e never really process two unrelated items consciously at exactly the same moment” (Dehaene, 2014). In other words, if we really do experience many dozens of distinct conscious stimuli simultaneously, as the intuition of richness might suggest, then consciousness is probably not a matter of high-level cognitive processes.

Of course, our intuition about richness could be radically mistaken: perhaps our experience is more limited than we think. Various debunking explanations have been proposed along these lines by late theorists, perhaps most famously the so-called “refrigerator light illusion” (O’Regan and Noe, 2001). The idea, in essence, is that we vastly overestimate the richness of conscious experience because whenever try to gauge whether we are having conscious experience of some thing or another, we thereby attend to and render conscious some perceptual state that was previously unconscious.[27]

Nonetheless, the intuition of richness is hard to shake, and arguably none of the various debunking strategies are wholly satisfactory. As noted by Peter Godfrey-Smith, “[m]uch of what Dehaene says... seems at odds with the simple idea that much of the time we experience a unified scene, with various things going on and eliciting our interest to different degrees” (2016). Consider also, for example, the experience of having an energetic and very engaging conversation while driving a car. It can certainly seem in such cases as though one’s conscious thoughts have been wholly occupied by the conversation, while the processes of navigating the traffic were carried out relatively autonomously without any conscious decision-making occurring. Nonetheless, it’s far from clear that in cases such as

[27] One might also attempt to chip away at the richness intuition via the notion of indeterminate or generic phenomenology. See Stazicker, 2011, and Grush, 2007.

this one's conscious perceptual experience was heavily degraded or largely absent, as the radically austere picture presented by Dehaene would seem to require.

*(ii) The limits of reportability*

I now wish to suggest a second and more controversial desideratum for a theory of consciousness, namely that it should be not too slavishly tied to report as a criterion of consciousness. This may seem a surprising suggestion, given that subjective report is still the normal measure of consciousness used in psychological work. For example, Papineau asks "what [is it that] makes a state 'uncontroversially unconscious' if it is not that subjects tell us so?", while Koch states that "if the subject denies any phenomenal experience, this should be accepted as a brute fact" (Papineau 2007, Koch 2007).

However, the idea that *all and only* conscious states are reportable has come under attack in recent years from several sources. One initial reason to doubt the constitutive connection between consciousness and report comes from growing worries among philosophers about the manipulability of subjects' reports about their experience, especially in cases where stimuli are masked or visible only for very brief durations. As discussed by Phillips (2015), in these cases, subjects' 'response criteria' can be influenced by all sorts of interventions that we would not expect to directly influence whether stimuli are consciously perceived. Notably, these include manipulating the prior probability that a stimulus will be presented and varying the monetary incentives for subjects. However, there are other myriad factors that influence such reports; as summarized by Phillips, these include "experimental instructions, task design, motivation, fatigue, and preconceptions about the experiment's purpose or intended outcome" (2015).

It is relevant, then, in light of the seemingly unreliable measures of subjective awareness commonly used in perceptual psychological that at least one set of experiments using masked targets with more subtle measures of awareness (Peters & Lau, 2015) found *no* dissociation between subjective and objective performance, even in tasks where conventional measures would suggest that subjects' performance was based on unconscious processing. As the authors put it, "subjects likely report lack of awareness because they interpret the response options in relative terms in the context of stimuli of various strengths."

Similar doubts about response criteria can also be raised in relation to failures of reportability in pathological cases such as neglect and blindsight (Weiskrantz, 1986). These cases have been widely taken to involve unconscious perception, insofar as subjects display perceptual sensitivity to the properties of presented objects yet deny having seen them. However, another possibility is that patients in these cases may still enjoy conscious perceptions that they are unable to report, perhaps because their phenomenology is degraded in some way or their mechanisms of self-monitoring have been impaired. Indeed, simply by varying the question asked to such patients, it is possible to elicit seemingly contradictory reports from subjects about their experience. For example, one study (Stoerig & Barth, 2001) discovered that a particular blindsight patient would give many more positive responses when asked "were you aware of something?" than they would if asked "Did you see anything?" Another result (Ramsøy and Overgaard, 2004) examined a blindsight subject who had previously denied having conscious experiences in her blind field when asked a simple yes or no question. They found that, when given the chance to report her degree of awareness using a four-point scale, she consistently reported some degree of awareness greater than zero whenever she had some objective sensitivity to the presence of a stimulus. As they put it, "her blindsight seemingly "disappeared" in the sense that... [a]ll correctness above chance seemed related to vague yet conscious vision.

Cases such as these are still highly contested (see, for example, Block 2016b, for some evidence of genuinely unconscious perception), but nonetheless give us some prima facie reason for seeking a theory of consciousness that either avoids reliance on reportability as a constitutive condition of consciousness or else explicitly justifies such reliance. [28]

Further pressure on the robustness of subjective measures of consciousness comes from 'no report' paradigms in neuroscience. Almost all neuroscientific experiments that Access theorists like Dehaene have used to link consciousness to specific neural activations rely on subjective reports as evidence that a subject saw a given stimulus. For example, in binocular rivalry experiments, subjects are typically presented with two conflicting images, one to each eye. In such experiments, subjects have the impression of alternating between conscious perceptions of each image, and required to indicate (often with a button press) when these shifts occur. Previous work (e.g., Lumer et al., 1998) had suggested that these conscious shifts are accompanied by greater activity in right fronto-parietal regions than in controls where subjects saw the same stimulus without rivalry (that is, with both eyes). This fact has been taken as evidence of the idea, for example, that frontal regions play an important role in engineering shifts in conscious awareness.

It has emerged that matters change somewhat, however, when binocular rivalry subjects are no longer required to report on their experiences. While their parietal and occipital areas continue to fire strongly, the strong frontal activations sometimes taken to be indicative of

[28] This is a particular problem for access accounts. However, it is worth noting that it may also threaten certain claims made by various Accessibility theorists. Thus Prinz (2011b) appeals to neglect patients as providing evidence of the role of attention in consciousness. However, once the reliability of subjective report as a criterion of consciousness is brought into question, it becomes an open question neglect patients lack the relevant conscious experiences.

shifting conscious awareness seem to be absent, even though other indicators (such as subjects' eye movements and subsequent reports) continue to suggest experience of alternating percepts. Summarizing this result, Tsuchiya et al. claim that these studies "point to the possibility that previously described switch-related activity in frontal cortices may be largely related to introspection and self-monitoring, associated with the difficulty of reporting ambiguous percepts, but not directly related to switching between conscious percepts" (Tsuchiya et al., 2015).

A final consideration that might tell against subjective reports as a robust measure of consciousness comes from a more theoretical consideration. In particular, one might doubt what independent reason we have for assuming that subjects are particularly reliable at assessing when they are having conscious experiences, especially in cases of neural deficits patients or where stimuli are masked or presented for brief intervals. There is, it has to be admitted, some intuitive appeal to the idea that we have privileged access to our own conscious states. However, it is clear in at least some cases that we are simply in error about this. A good summary of these kinds of cases is provided by Eric Schwitzgebel (2008). One such simple example comes from failures of emotional self-awareness: most of will be familiar with cases where we have been feeling angry or jealous, yet seemingly sincerely denied feeling any such emotion, at least initially.[29] Framed as a challenge to the access theorist, then, we might ask what independent reason we have to assume that subjects' introspective judgments about whether or not they underwent a given experience are reliable.

To summarize, then, I would suggest a theory of consciousness should not be too

[29] An anecdote from a colleague illustrates this point well. Visiting her therapist one day, she was asked how she was feeling, to which she replied, quite reasonably, that that was precisely what she was paying the therapist to find out.

closely tied to reportability as a criterion for consciousness; or, if a close link between consciousness and subjective report is indeed to be defended, substantive arguments must be advanced in this regard. While a few theorists (notably Rosenthal, 1997) have attempted to defend the importance of reportability as a criterion of consciousness, the majority of access accounts have little to say in this regard. Dehaene, for example, states simply that "the notion of a phenomenal consciousness that is distinct from conscious access is highly misleading and leads down a slippery slope to dualism." This kind of consideration is unsatisfactory, however, in light of the complex practical and theoretical issues involved in relating consciousness to report.

*(iii) Constraint*

A third desideratum for a theory of consciousness should be that it offers us some principled way of *constraining* the range of states that are conscious in any given case. The human nervous system contains numerous regions that engage in quite complex information processing tasks yet fail to be conscious. For example, the enteric nervous system regulates a vast number of physiological functions and closely interfaces with the brain, yet the majority of its behavior (such as the selective release of cholecystokinin in response to intestinal pathogens) clearly occurs without consciousness (Mayer, 2011).

Similar points can be raised for activity in the LGN, the cerebellum, and the spinal cord. Thus, Tononi & Koch suggest that a theory of consciousness must explain why "the cerebral cortex gives[s] rise to consciousness but the cerebellum does not, though it has even more neurons and appears to be just as complicated" (2015). Indeed, many of these areas can seemingly contribute to whole organism behavior, such as learning to avoid damaging stimuli. For example, an astonishing study by Liu et al. (2005) show that a transected chunk of rat spinal column was able to learn (via standard instrumental conditioning techniques) to selectively hold a hindlimb in a particular position. However, such behaviors are presumably carried out entirely without consciousness.

Thus another desideratum for a theory of consciousness should be to explain *why* these kinds of information processing fail to be conscious, and to set a limit to the range of states that can be conscious.[30] In doing so, there is at least some reason to think that this limit should not be set *too* broadly. The concern I have in mind here is that a key part of why we are interested in consciousness to begin with is its connection to ethical and practical concerns that are cashed out *at the personal level*. The reason we opt for anesthetics in painful surgical procedures, for example, is that we trust they prevent *us* from undergoing unpleasant conscious experiences, and the reason we are keen to assess the presence of consciousness in a relative in persistent vegetative state is that we have an interest in whether the *person* we care about still has a subjective window on the world. In neither case would it be of much relevance to us to be informed that isolated subjective experiences were still taking place in the enteric nervous system or spinal cord, unless these translated to experiences that were in some way present in the subjective experience of the whole person.

I do not wish to put too much weight on this point, and of course, a science of consciousness has to be open to surprising conclusions. However, the goal of a theory of consciousness is to make intelligible to us how brain processes give rise to *our* experience,

[30] Of course, one might satisfy this desideratum by claiming that consciousness is, in fact, ubiquitous, or at least present in very simple systems we would not normally think of as conscious. This is the approach taken, for example, by Tononi & Koch (2015).

and it will be better able to do so to the extent that it can be clearly related to the kind of experiences that we undergo as individuals.[31]

(iv) *Integration*

A final desideratum for a theory of consciousness is that it should offer at least the beginnings of an account of how different types of conscious state can all be integrated into a single subject's awareness as intelligible and meaningful states.[32] Thus at a given time, I may be undergoing a conscious auditory experience of a seagull's cries and a visual experience of a setting sun. At the same moment, I might also be consciously reflecting on its beauty, and calling to mind memories of previous sunsets I have seen. Here, we have several different kinds of mental state – auditory and visual perceptions, thoughts, and images – that nonetheless jointly constitute my total experience at a time and can rationally inform my behavior. The challenge, then, is how these quite disparate phenomenal states are all jointly intelligible to me as a subject.

Theories such as Dehaene's Global Neuronal Workspace or Rosenthal's higher-order thought view which take all conscious experience to depend on a single mechanism have a natural starting point for answering this problem. Since they take the distinct experiences that contribute to my awareness to be constituted by the operations of a single process (such as activity in the Global Neuronal Workspace) with distinct psychological

[31] Another way to make this claim would be to imagine that we accept a version of panpsychism as true. In that case, there would be a further question about which of the subjective experiences generated by constituent atoms are *mine.* This is not simply a version of the notorious combination problem (James, 1890), which is concerned with the metaphysical problem of how different experiences might be combined, but rather the suggestion that in addition to giving a theory of subjectivity, we are also required to give an account of the boundaries of personal-level experience.

[32] I take this problem to be somewhat related to the issues relating to the unity of consciousness (Bayne, 2010). However, even if we had an account that explained how conscious experiences came to be phenomenally unified, we would still need some explanation of how these different phenomenal states were all similarly meaningful to a subject despite apparent differences in their sensory qualities and representational formats.

characteristics, my capacity to integrate my conscious states with one another can be explained in terms of the operations of this process.[33] The same is true of working memory accounts of consciousness, like those endorsed by Baddeley and Cowan. Accessibility accounts can also point to a single psychological property that underwrites the integration of conscious experience, such as availability to working memory (Prinz, 2012) or to higher-order thought (Carruthers, 2005). While these may not be completely satisfactory accounts of phenomenal integration, depending as they do on a dispositional rather than occurrent property, they can at least make intelligible how a perceptual state that is not cognitively accessed may nonetheless be psychologically and phenomenally integrated with my other states.

The theories that are most threatened by this sort of challenge are access-independent theories. Thus Block (2007) and Lamme (2010) suggest that perceptual consciousness may consist in certain kinds of neural process in sensory areas, and that such states may be conscious without being cognitively accessed or even accessible to the subject. This raises the question as to how those states are able to contribute to a subject's integrated phenomenal awareness at all. Of course, these theorists can appeal to biological properties (such as, for example, 40Hz oscillations) in answering the question. However, this is not entirely satisfactory, and leaves us questioning why this neurobiological relation rather *that* one should suffice to make it the case that a cognitively inaccessible state nonetheless contributes to my awareness of the world.

## 4.4 – CSTM and the Workspace-Plus model of consciousness

[33] One complication here is that Dehaene sometimes seems to have in mind the view that we can only attend to a single thing at a time. Elsewhere, however, he has suggested that the contents of experience are constituted by the active contents of working memory (Dehaene, 2016; see also King, Pescetelli, & Dehane, 2016).

Some of above desiderata may be seen as controversial, and they are not meant to be exhaustive. Likewise, I would not suggest that any of them are absolute requirements for a theory of consciousness. Just as many theorems of quantum mechanics are only comprehensible in mathematical terms, it may be that the correct theory of consciousness will take the form of a detailed neurobiological model that bears no relation to ordinary platitudes about consciousness. However, part of the task of a theory of consciousness is to create a bridge of sorts between scientific and pretheoretical conceptions of consciousness. On one side of the equation stand desiderata like those above, which reflect our phenomenological and intuitive ideas about consciousness, while on the other side we have facts about neural and psychological processes.[34] If we are comparing theories of consciousness, then, it is reasonable to ask to what extent a given account makes intelligible a connection between the two.

With this in mind, I now return to what I earlier described as the Workspace-Plus model of consciousness in which I take CSTM to play a central role. The starting point for this approach is the claim that while the kinds of central cognitive processes described by Dehaene and others may constitute an important part of conscious experience, they are not exhaustive: there is something they are leaving out. The next claim is that what is being left out is not just a matter of sensory representations of the kind favored by Lamme, but rather something more closely connected to an individual's awareness of the world in all its significance: a rich panoply of fleeting but *meaningful* representations of objects and events in the body and the environment. The final claim is that the current best candidate

[34] Cf. Godfrey-Smith (2004): "Philosophy should aim to describe the connections between facts about the use of difficult and controversial concepts, and facts about the parts of the world that the concepts are in some sense aimed at dealing with."

mechanism for this mode of awareness is encoding in CSTM, in virtue of its large capacity, brief duration, and capacity for semantic encoding.

I hope that this sketch of the model has some prima facie plausibility, and is at least worth exploring. There are various ways to develop the Workspace-Plus model, as I will shortly examine, but before that, it is worth spelling out some of the central features that will be shared among different versions of the account. The first is that it characterizes the key mechanisms of perceptual experience neither with strictly perceptual processes (as do Lamme and Block) or with central cognitive processes like working memory or global broadcast (along the lines suggested by Dehaene, Kouider, Cowan, and others), but rather an intermediate set of processes that we could broadly term *postperceptual* (see Chapter 5).

The second is that the theory claims that perceptual experience has a unifying format (cf. Baddeley's episodic buffer). This serves to further distinguish the account from theories like that of Block and Lamme that allow for strictly sensory forms of consciousness. I should stress, however, that the Workspace-Plus model is not committed to idea that there are *no* differences in format among the various contents of perceptual experience, but merely to the claim that there is also *some* respect in which they are encoded in the same format. This accounts for the way in which representations from different sensory modalities can all contribute to a uniformly intelligible awareness of disparate events in the body and the environment, and for our ability to immediately respond to quite different perceptual experiences not in respect of their sensory qualities but in terms of how they are relevant to our interests and goals.

The third key feature is that, unlike access theories, it distinguishes between two basic mechanisms of awareness, one a central cognitive process like the global neuronal workspace, and the other a short-term buffer like CSTM. Whereas the former mechanism specializes in more focused demanding processing of the kind involved in conscious reflection or focal perceptual attention, the latter underlies our fluid and effortless perceptual awareness and prereflective understanding of our immediate environment.

These, I take it, are the shared core commitments of the Workspace-Plus view. However, there are many more specific ways the theory can be spelled out. I take it that the single most problematic question concerns the *format* of representations in CSTM (see Fig. 4a). I have suggested (in Chapter 4) that information encoded in CSTM may constitute a kind of *hybrid* representation, consisting of a conceptual representation element bound to a sensory representation. However, even granting this, there is potential disagreement about the syntactic structure of the conceptual element in such hybrid states. One approach mentioned in the previous chapter would be to claim that this structure is somewhat simpler than a proposition, consisting of a concept and a strictly referring element (GW+$_1$; see the discussion of Burge in Chapter 5). Alternatively, one might argue that it has a fully propositional syntax (GW+$_2$).

Another variation of the view might claim that the high-level semantic encoding in CSTM is accomplished by *some non-conceptual* semantic representation bound to a lower level sensory representation (GW+$_3$). This version of the view might, for example, be broadly amenable to thinkers such as Burge (2010b), Tye (1995), and Siegel (2010), who may wish to allow for high-level nonconceptual representations. A final option might be to insist that representations in CSTM contribute directly to our perceptual experience only in virtue of their *conceptual* content rather than their sensory content (GW+$_4$). This last strategy may be appealing, for example, for theorists (such as McDowell, 1994, or Brewer, 2005) who are independently motivated by the arguments in favor of Conceptualism.

The different options regarding format allow for a variety of positions within the broad framework of the Workspace-Plus model. However, I take it that they all share the core commitments given earlier. Moreover, for all of these approaches, CSTM seems like a natural candidate mechanism for encoding the relevant contents that constitute perceptual experience.

| | View |
|---|---|
| Hybrid sensory & conceptual;<br>non-propositional | **GW+$_1$** |
| Hybrid sensory & conceptual;<br>propositional | **GW+$_2$** |
| Hybrid sensory and semantic-nonconceptual;<br>non-propositional. | **GW+$_3$** |
| Strictly conceptual;<br>propositional | **GW+$_4$** |

**Fig. 4a –** Varieties of Workspace-Plus account

A second kind of decision concerns whether the contents of CSTM are necessarily accessible to Central Cognition, or whether there might be cases (such as, perhaps, blindsight and neglect) in which representations are conscious despite being cognitively inaccessible. All of the above variations may be spelled out in ways that allow for inaccessible conscious in CSTM, or hold that *only* those states of CSTM that are accessible to central cognitive constitute my perceptual experience. While I am open to the idea that representations in CSTM may not always be cognitively accessible, theorists such as Prinz (2012), Carruthers (2005), and Tye (1995) who take cognitive accessibility to be a necessary condition of a state's being conscious may favor the latter approach.

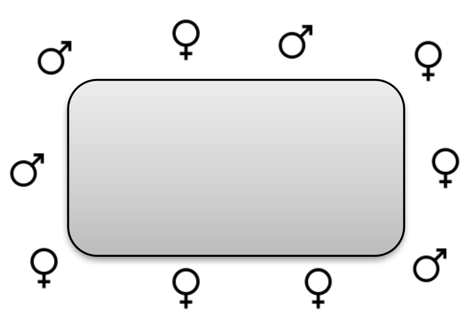
There are also questions concerning the role of Central Cognition in my account. I have suggested that Central Cognition may underpin a distinctive kind of awareness, such as that involved in reflective thought and attentionally-demanding perceptual activities, such as finding a friend in a crowded supermarket, or singling out one voice from a hubbub. However, I wish to remain broadly neutral on other questions about how Central Cognition contributes to experience, such as the question of whether we undergo conscious experience without any sensory component (Carruthers, 2015a; Prinz, 2012).

I hope that the variety of different implementations of the theory given above may make it an attractive option even for theorists who are skeptical of other specific claims about format and accessibility. I will shortly spell out what I take to be the chief attractions of the broad framework. However, moving forward, it would also be helpful for me to spell out which specific implementation of the Workspace-Plus theory I take to be most promising. To that end, I can say now that the version of the view I find most attractive takes encoding in CSTM to involve hybrid conceptual-sensory representations with a non-propositional syntax that are not constitutively accessible to Central Cognition ($GW+_1$ in the above figure).

What might be the advantages of this specific account over other versions of the view? There are considerations that push me towards this particular implementation, as I will now describe. However, it is important to note that the following arguments are tentative, and presented as much to account for my preference for this form of the theory than to convince steadfast opponents.

First, I think an adequate account of the character of perceptual experience should make reference to the distinctive phenomenological contributions made by the sensory and categorical elements involved in awareness of a given object, and this is best accomplished via the kind of hybrid conceptual-sensory format adopted by $GW+_1$. These two elements, I suggest, seem to be manifest at the phenomenological level: looking at a tawny lion, for example, the phenomenal character of my experience includes both strictly sensory phenomenology relating to its color and shape, but also a distinct and more elusive kind awareness involved in my experience of it *as* a lion.[35] The hybrid nature of states of perceptual experience, I would suggest, also presents itself in the different ways in which we can make *use* of representations in experience (Kosslyn, 1994). Thus, looking at an object, I can recognize that it is *a floor rug*, and that it has a picture of the Eiffel Tower on it, while also considering its sensory elements such as its color and shape to gauge how it would look inside my apartment (and whether it would tie the room together).

I do not take the notion of such hybrid formats to be conceptually problematic; I think we frequently make use of such representations in daily life in the forms of maps and charts (Camp, 2007). As a simple example, consider the seating chart shown below (Fig. 4b). Here, I can exploit the strictly symbolic elements of the representation to conclude, for example, that there are more women than men around the table. However, I can also exploit the analogue spatial structure of the chart to derive a very large number of more specific conclusions, for example that a man in the top-left hand corner is seated diagonally opposite another man, that the man beside him is sitting slightly further back, and so on.



[35] As noted in the previous chapter, I am open to various approaches for spelling out the precise phenomenological contribution of categorical perception.

**Fig. 4b –** A simple example of a hybrid representation showing men and women seated at various point around a table.

One tricky question for such a hybrid representation account concerns *neural vehicles*, and specifically how the binding mechanism between sensory and conceptual elements might work. For example, one possibility is that sensory information is literally present in the neural mechanisms that underpin CSTM. Another possibility is that CSTM serves to 'index' or otherwise 'point to' rich sensory representations in lower level sensory areas, perhaps rendering them conscious via some mechanism of neural feedback (cf. the "joint determination" view considered by Lau, 2008, and Lau & Brown, forthcoming). These questions, I would suggest, are best left for cognitive neuroscience to answer.

Why could the semantic component of these bindings not be accomplished by some form of nonconceptual representation? A proper answer to this question would require giving a theory of concepts, something that lies beyond the remit of this dissertation. However, briefly, my reasons for taking the relevant semantic content to be conceptual rather than nonconceptual stem from two considerations. The first is that I am independently sympathetic to relatively undemanding neo-empiricist accounts of concepts that, broadly speaking, reject strict epistemic requirements on concept possession, explaining it instead in terms of the acquisition via perceptual experience of reusable templates or proxytypes (Prinz, 2002).[36] As a result, I am inclined to regard concept learning and deployment as a relatively unsophisticated process intimately linked to perceptual learning, as opposed to being constitutively tied to more complex psychological operations such as logical inference. The second is that, for the kind of reasons described in the previous chapter, I remain

[36] For the kind of account I have in mind, see also Lyrra (2005): "the conceptual sphere could be thought of as differing from the nonconceptual sphere in that non-conceptual content makes up fine-grained perception-like representational content and conceptual content abstracts some fine-grained nonperception-like higher-order features from the non-conceptual sphere."

unconvinced that any purely sensory or pictorial representation could adequately account for the kind of high-level properties that I take to be present in perceptual experience: I struggle to see how a picture, a map, or a chart could unambiguously (and without recourse to conceptual content) represent properties like ‘being drunk’, ‘being a CD player’, or ‘being a predator’.[37]

Finally, one might wonder why I take representations in CSTM to lack a propositional syntax. Note that it is important to distinguish questions about syntax from questions about content. When we ask whether a given representation has a propositional syntax, we are asking about the structure of that representation as realized in the mind of a subject, and whether, for example, it can be decomposed into constituent representational parts. This may in turn be important if we are interested, for example, in the range of concepts that a creature can be said to possess (Heck, 2000). By contrast, if we ask instead whether a given representation has propositional content, we are concerned with the broader question of whether that representation has truth or veridicality conditions that can be spelled out in terms of a particular set of propositions, at least in principle independent of concerns about the internal structure of that representation.[38] The distinction between the two questions allows for the possibility that some representations may have propositional content but not propositional syntax (this may be the view of Siegel, 2010).

Confining ourselves to questions about propositional syntax, then, I will stress that the

[37] It might seem as though there is a tension here between these two points: if sensory representations cannot represent high level contents, how could concepts *built out of* sensory representations do any better? Of course, there is a lot to be said here (see Prinz, 2002, for a fuller discussion). However, in short, I think the two points in fact complement each other: a well-developed account of how sensory representations could form interconnected complexes that could allow for category identification and semantic relations might easily qualify, in my terms, as a theory of concepts.

[38] Cf. Burge “Some have thought that whether perception is propositional depends on whether what is represented is an object or a state of affairs. I believe that this view has things backwards. A state of affairs can be represented either with a singular representational structure or with a propositional structure.” (2010a)

claim that encoding in CSTM lacks full-blown propositional structure is the aspect of $GW_1$+ that I am least committed to. Nonetheless, especially in light of my neo-empiricist sympathies mentioned earlier, I see it as a viable hypothesis that concepts could be tokened in perception and perhaps constitute states with propositional content without being part of a representation with a complete propositional structure. To give a toy example, imagine that an individual encodes a sensory state in CSTM in respect of the concept BLUE. The best way to spell out the veridicality conditions of this state might be in terms of the truth or falsity of some proposition; for example, 'there is at least one blue object in location L and that object non-deviantly caused this state', or something similar. This might be the case even if the structure of that representation was fairly simple, consisting, perhaps, of a demonstrative referring element and a concept functioning attributively, as in 'That-BLUE'.

This is just a sketch of a proposal, but I do wish to remain open to the idea that the kind of categorical perception that I suggest is performed by CSTM involves the application of concepts, even if representations in CSTM have a very simple structure. If this proposal is along the right lines, it might even explain categorical perception in the case of animals that lack higher cognitive capabilities. This is a consideration that I will briefly return to in Chapter 5.

## 4.5 – Advantages of the Workspace-Plus Account

Now that I have given some considerations in support of $GW+_1$ as a specific implementation of the Workspace-Plus model, I can now turn to a broader evaluation of the theory. What might be the attractions of the Workspace-Plus approach? In particular, what might it offer over and above existing accounts?

For one, it seems to be better placed than many other accounts to satisfy the desiderata listed earlier. First, it seems better placed to accommodate the criterion of phenomenological adequacy than the leading access-only accounts, specifically in its treatment of richness. One of the least appealing features of such theories, as noted above, is their austere portrayal of perceptual experience. However, supplemented with a high-capacity short term buffer in the form of CSTM, such models could better accommodate the apparently rich and fleeting nature of human experience without recourse to debunking explanations. When I stand in Times Square, for example, I may deploy Central Cognition in noticing and attending to some items (perhaps by encoding them in working memory), while still enjoying rich but fleeting experiences constituted by the contents of CSTM.

This kind of 'two factor' account of consciousness can also accommodate the data from partial report paradigms. Thus consider the apparent divergence between what subjects can report and what they claim to see in partial report cases like the Sperling Test. Whereas the access theorist must appeal to various debunking strategies (such as generic phenomenology) to explain these, a Workspace-Plus account can take the data at face value, allowing for the possibility (like access-independent theories) that subjects really do consciously see all of the items in the initial array in respect of their specific alphanumeric identities, while being nonetheless unable to report on all of them. Thus even though the CSTM theorist agrees with the access theorist that sensory memory is not by itself conscious, she can nonetheless allow that, in virtue of the rich capacity of CSTM, subjects really did have fleeting specific phenomenology of all the relevant letters and rectangles, including their specific semantic identity.[39]

[39] If we appeal to CSTM to explain subjects' phenomenology in the Sperling Test, as suggested here, we do face one further puzzle for interpretation of earlier work on iconic memory: why, if there is rapid conceptual processing of items presented in an array, did subjects in Experiment 6 is Sperling's original paper fail to exhibit partial report superiority when cued to report on items in respect of their semantic identity (e.g., "letter" or "number")? There are

A second advantage of a Workspace-Plus model as compared to access theories is that can accommodate some cleavage between consciousness and reportability. By distinguishing between the mechanisms of central cognitive access and a short-term conscious buffer, a Workspace-Plus has additional theoretical flexibility, and can allow for the possibility (which I take to be empirically open) that, for example, some blindsight subjects might lack cognitive access to their conscious states.

In this regard, the Workspace-Plus model might also be able to cast light on the findings of no-report paradigms. It may be the case, for example, that in cases where subjects are not required to report on presented items, they do not actively encode presented items in conscious working memory, despite nonetheless briefly experiencing them in CSTM. This could explain the different sorts of neural activity observed in no-report cases.

Of particular note in this regard is that the Workspace-Plus model can allow for such failures of report while also satisfying the desideratum that a theory of consciousness be able to *constrain* the range of states that are candidates for consciousness. In particular, sensory representations would not be candidates for consciousness prior to encoding in CSTM. This would seemingly exclude the possibility of, for example, wholly non-integrated conscious states arising just in early visual areas, and would provide us with a way of ruling out isolated instances of consciousness in the spinal cord or the Enteric Nervous System. Thus the kind of worries about consciousness in isolated brain regions or spinal mooted by Prinz (2007) in relation to Block's account could therefore be averted.

Relatedly, a Workspace-Plus model can at least begin to sketch how our different

several interesting possible explanations for this, but I will mention one possibility, namely that subjects may have encoded the items in the array in respect of more specific semantic categories ("r" or "7") than those cued by Sperling ("letter" and "number"). Thus the number of characters that were reportable would be limited by subjects' ability to make rapid inferences about the determinable class the individual items belonged to. This would be a valuable area for further experimental work.

conscious states might satisfy the *integration* desideratum suggested above. To see how this might work, note that, ex hypothesi, mental states in both CSTM and Central Cognition are all encoded at least partly in respect of semantic categories. It therefore suggests a mechanism of integration among incoming sensory information and a subject's cognitive systems, namely a unified semantic format. This puts the theory in a promising position to explain some puzzles about the meaningfulness of our perceptual experience, and the apparently easy flow of information between conscious perception and conscious thought: all of these states would share something in common, namely being partly encoded in a single format.

As a final point in its favor, note that the theory is directly open to empirical falsification: if it is demonstrated that neither CSTM or any equivalent semantic buffer actually exist, the theory cannot be right.[40] While I believe, as argued in Chapter 2, that there is strong data in support of CSTM, it has still been subject to comparatively little investigation as compared to working memory and sensory memory. There has also been, to my knowledge, almost no neurobiological investigation to suggest how it might be implemented in the brain. Nonetheless, I think the account of consciousness sketched so far offers a promising new avenue of investigation. Of course, even if CSTM can be decisively shown to exist, that would not immediately validate the Workspace-Plus model of consciousness, but it would put the theory on a much firmer empirical ground.

## 4.6 – Evaluating the Workspace-Plus Model

[40] One experiment that would be particularly helpful in this regard, as noted in Chapter 2, would be to apply the masking techniques adopted by Maguire & Howe (2015) with Potter et al.'s (2014) original 12-item post-stimulus cueing methods.

The above discussion hopefully serves to show how a Workspace-Plus model might work, and some of the points in its favor. I now wish to briefly consider and rebut some objections to the view as outlined above.

*(i) Does all perceptual experience involve categorical perception?*

The single most important challenge for the Workspace-Plus model concerns the relation between perceptual experience and categorical perception. I argued in the previous chapter that the basic function of CSTM was perceiving-as, and I have argued in this chapter that the contents of CSTM constitute the contents of perceptual experience. It would seem to follow, then, that any perceptual experience involves perceiving some *X* as F, where 'F' indicates some categorical property or properties already possessed by the subject.

While I do not think the Workspace-Plus theorist need necessarily be committed to this in absolutely *every* case (see below), I do think it is a natural and independently attractive conclusion for the vast majority of cases of perceptual experience. Indeed, I take it as a positive virtue of my account that it can appeal to this sort of semantic encoding to explain why our perceptual experience is immediately comprehensible to us in terms of its relevance to our goals and motivations: in conscious perception, I suggest, we do not merely undergo sensory experiences of objects of various colors and shapes whose significance we can subsequently ascertain, but rather become immediately aware of an object *as* an apple, or *as* a snarling tiger.

Moreover, while I do not rest anything on the point, it is also perhaps worth noting that in claiming that perceptual experience constitutively involves that application of concepts (or nonconceptual semantic categories, in the $GW+_3$ version of the view), I am not being wholly perverse, but am in some good philosophical company. Thus Jerry Fodor, expressing sentiments I broadly share, writes as follows.[41]

> On the one hand, there is *no perceiving* without *perceiving a*s; and, on the other, there is no *perceiving as* without conceptualization; and, on the third hand, there is no conceptualization without conceptual content. (Fodor, 2015)

Likewise, recall the quotation from William James in the previous chapter.

> [W]e never get an experience that remains for us completely nondescript: it always reminds of something similar in quality, or of some context that might have surrounded it before, and which it now in some way suggests. (James, 1900)

However, I must grant that there are some cases of perceptual experience in which categorical perception seems to fail us. Consider the experience of waking up in an unfamiliar environment and seeing some mysterious object above one's head. In such a case, we might puzzle for a few moments to work out that what we are seeing is, for example, an unusually shaped light fixture. Another example comes from the weird experiences of colors and patterns that accompany intense hallucinations. Likewise, consider the perceptual experiences undergone by associative agnosics: despite lacking the ability to classify what they are seeing, they nonetheless undergo conscious experience, as demonstrated by their facility in copying images. Are these not cases of perceptual experience without categorical perception?

While I admit that these examples illustrate cases in which categorical perception may barely get off the ground, it is not clear to me that it is wholly absent. While the role of

[41] See also Burge (2010a): "Perceptual states must represent what is perceived as being a certain way."

CSTM is to encode perceptual inputs in respect of stored categories, it may be that, when confronted with an unfamiliar object or sensation, it struggles to do so in terms of any existing high-level category. However, this is compatible with the idea that CSTM encodes these experiences in terms of some *basic* category or another, even if that is a highly indeterminate one (such as “that spikey green thing”, or “that moving object”). In this regard, it is important to note that even associative agnosics can recognize and name very simple geometric shapes (see Gottfried, 2011: Ch.10), suggesting that they have at least some preserved ability to categorize what they are seeing in at least very basic terms.

However, this account will only work if there are at least *some* categories I can apply in perception. Could there be an experience that lacked any kind of categorical content at all, not even being identifiable in respect of basic color or shape concepts? Again, perhaps some of the strange feelings reported during psychedelic experiences might be candidates in this regard. One option for the CSTM theorist here might simply be to allow that sometimes CSTM draws a blank, and cannot classify an experience in respect of any relevant existing concept. This might seem a fatal concession for the CSTM theorist, but note that in granting this point, we need not grant that perceptual experience occurs without conceptualization. In particular, it seems open to the CSTM theorist in dealing with these cases to suggest that, when confronted with extremely novel sensations that defy categorization in terms of existing concepts, CSTM may respond by generating a *new* label or category-marker for the relevant state. Though encoding in respect of this category would not initially contribute to the *understanding* of the experience in which it features, it might nonetheless be re-applied to similar states in future and allow for recognition. After all, even in cases of bizarre hallucinations or strange drug-induced sensations, people are often able to recognize their feelings as being similar to ones experienced during previous psychedelic episodes.

This kind of proposal might also gesture towards possible ways in which the CSTM account could be developed to account for concept learning in neonates, or other cases (such as, perhaps, recovery from congenital blindness) in which an individual is forced to undergo a rapid process of perceptually-directed learning. While it is an open possibility that infants begin with a stock of some basic innate concepts (Carey, 2009), it also seems likely that many sensory states they undergo will be entirely novel. Starting from scratch, as it were, and building new initially uninformative perceptual categories on the fly, such an infant's experience might well be described, in James's familiar terms, as a "buzzing, blooming, confusion" (1890).

*(ii) Is CSTM truly multimodal?*

A second objection to my view stems from the questions raised in Chapter 2 concerning whether CSTM operates as a single semantic buffer across different sensory modalities, or, alternatively, whether distinct senses such as a vision and audition may each have proprietary forms of CSTM. As noted there, I regard this as an open empirical question.

If CSTM is indeed a single faculty serving all modalities, then no issues arise for the theory as presented. However, if it turns out that different sensory modalities all have their own semantic buffers, one might worry that this compromises some of the attractions of the view. For example, if consciousness truly operates across different 'channels' tied to specific sensory modalities, one might worry that it undermines some of the appeal of the theory concerning the integration of phenomenally conscious states.

While matters for me would certainly be simpler if CSTM turns out to a truly unitary mechanism, I think the theory can weather the alternate possibility. Note that my account explains our capacity to integrate different phenomenally conscious sensations not primarily in terms of realization by a single mechanism, but in terms of conscious states' possession of a single unified conceptual format. Thus, if it turns out we really do have a plethora of CSTM-like buffers, we can still explain how it is that they can all immediately inform a subject's behavior in respect of relevance to their goals, namely by virtue of their shared format.

Nonetheless, there are two ways in which a 'multiple buffer' story might seriously trouble my account. The first is that it may seem to make the phenomenal unity of conscious experience a bit mysterious: why should these different buffers all contribute to a single 'phenomenal field'? However, I take it that are a number of viable options for the Workspace-Plus theorist in settling this question, including, for example, Bayne's mereological account of phenomenal unity (Bayne, 2010). A more directly threatening objection would arise if it was discovered that some forms of conscious experience – bodily sensations, or pain, perhaps – were processed in a way that did not involve any mechanism even vaguely analogous to CSTM. At this point, the Workspace-Plus might attempt to salvage the theory by confining its explanandum just to those conscious states (such as, perhaps, visual and auditory experiences) that are indeed processed in semantic buffers. However, this is not a particularly theoretically attractive move. Much of the appeal of the broader approach, then, rests on the empirical question of whether semantic processing of the kind performed by CSTM is based in a single multimodal mechanism; or, failing that, whether some form of semantic encoding is at least operative across all of the major domains of conscious experience.

*(iii) Aren't some representations in CSTM simply unconscious?*

Another objection concerns the claim made by the Workspace-Plus model that encoding in CSTM suffices for conscious experience. As noted in Chapter 2, some stimuli encoded in CSTM (such as those involved in attentional blink effects) are unreportable. Moreover, some of the stimuli used by Potter were presented for extremely short durations – as little as 13ms. Is it really plausible that subjects consciously perceived these stimuli?

I think we should certainly be open-minded about the possibility that subjects in these experiments underwent conscious experiences despite being unable to report them. As argued in 4.3, above, a number of different theoretical and empirical projects are lending weight to the idea that report is an imperfect measure of conscious experience. I thus regard it as a live empirical possibility that subjects in, for example, attentional blink paradigms consciously see stimuli despite claiming they saw nothing. It is also worth repeating the idea mooted in Chapter 2 that, in light of the work of Maguire & Howe (2015), it is possible that the stimuli presented by Potter et al. (2014) for the very briefest durations of 13ms and 33ms were inadequately masked and may been encoded in some sensory form of memory such as fVSTM rather than CSTM. However, the fact that Maguire & Howe (2015) found that masking had no effect on stimuli presented for 50ms or more suggests that these stimuli were not encoded in sensory buffers. For a theorist who was otherwise sympathetic to a Workspace-Plus account but who balked specifically at the idea that conscious perception could occur for very short presentation times of 13ms, this may be an important distinction.[42]

Finally, it is worth noting that, if we wished to preserve a link between reportability and consciousness, we could modify the Workspace-Plus account so as to specifically require that *only* those contents of CSTM that were cognitively accessible contributed to conscious experience. This would then make the account a type of accessibility theory. While I take this

[42] Thanks to Eric Mandelbaum for pressing me on this point.

to be a viable theoretical option, I think it is worth exploring first the idea that the contents of CSTM are all conscious, since this would allow us to provide a theory of consciousness that was couched exclusively in terms of strictly occurrent rather than dispositional processes.

*(iv) Mental latex*

Another objection to the view might question whether there may be wholly non-representational features present in perception, or in, Block's terms "Mental Latex" (Block, 1996). For example, Block suggests that "the phenomenal character of the experience of orgasm is partly non-representational", and that in "auditory and visual represention s as of something moving, there are experiential but non-representational differences between these representations in different sensory modalities."

This might seem to present a prima facie challenge to my view, insofar as I take a representation's being encoded in respect of some semantic category or another to be a basic feature of experience. Again, however, this possibility need not be in conflict with the view I have described. For one, note that the specific version of Workspace-Plus model I am defending (GW+$_1$) endorses a hybrid account of representational content in which experience have both conceptual and sensory (iconic) features. It is an open possibility, then, that the phenomenal character associated with certain *sensory* features may play no representational role, provided that these representations are *also* encoded in respect of some semantic category or another.

However, Block also suggests that some experiences may lack representational content *tout court*, giving, as noted above, the example of the experience of orgasm. This sort of objection has already been considered earlier, but in relation to this specific example, it is far from clear to me that orgasms constitute a case of experiences that lack any representational properties whatsoever. A natural way of coarsely describing the character of an orgasm is as a pleasurable event located somewhere in our bodies.

It might be objected that 'being pleasant' is not a semantic category. However, one important point to note is that I am very open to the idea that the content of some forms of categorical perception might be cashed out in non-indicative terms. I think it independently very likely that the best account of the content of certain representations in CSTM may appeal to imperatival contents or 'experienced mandates' (Martinez, 2011; Siegel, 2014; Klein, 2015). In fact, I believe that "pleasant" and "unpleasant" may be some of the important and evolutionarily ancient forms of category that are applied in conscious experience. However, this is a suggestion that I will return to in Chapter 5.

*(v) Unconscious semantic processing?*

A fifth objection concerns whether the Workspace-Plus model is committed to the idea that *all* semantic processing is conscious. In particular, whatever definition of semantic processing we adopt, it is likely that a lot of sophisticated but uncontroversially unconscious processing involved in, for example, language comprehension will likely qualify as semantic.

This is an important point for the Workspace-Plus model to recognize, but there is no reason to assume that all semantic processing must be routed through CSTM. In particular it seems an open possibility that language production and understanding is subserved by dedicated psychological mechanisms, given the relative frequency of neural deficits (such as receptive and expressive aphasias) that specifically disrupt grammatical or lexical understanding and production. More generally, the Workspace-Plus theorist can remain open to be possibility that there are forms of semantic processing in the brain that occur outside CSTM, and which do not directly contribute to conscious experience.

*(vi) Animals*

A final and very important objection concerns the question of consciousness in non-human animals. If we assume, as many philosophers have, that animals do not possess concepts, then it might seem as though they must also lack any analogue of CSTM. However, it seems an open question whether such animals may nonetheless undergo perceptual experience.

Of course, one option for the CSTM theorist would be to simply bite the bullet and insist that most animals are not in fact conscious. However, this is not the kind of approach I wish to endorse. Rather, I am inclined to think that a large number of non-human animals possess concepts, including, perhaps, even simpler creatures like honeybees (Carruthers, 2009). Of course, another option for the CSTM theorist would be to adopt a theory like $GW+_3$ that explains semantic encoding in nonconceptual terms. I will return to considerations of issues such as these in the next chapter.

## 4.7 – Rethinking cognitive access?

As a final exercise, it is useful to consider the place of the Workspace-Plus model in the broader theories of consciousness debate. In particular, we might ask whether it can be located in the tripartite scheme discussed earlier. I have already explained some of the major differences and advantages I believe the theory to possess relative to access theories. However, the Workspace-Plus theory equally does not seem to be an accessibility account of the kind offered by Prinz (2012) and Carruthers (2005). As noted earlier, Prinz and Carruthers both appeal to dispositional properties – availability to working memory or to higher-order thought – to explain why some states are conscious and others are not. By contrast, a Workspace-Plus account explain consciousness in terms of occurrent processing in a specific psychological mechanism, namely encoding in working memory and CSTM.

Does this show that the theory is a novel kind of access-independent theory? This seems the natural place to locate the view, but before doing so, it is useful to examine a recent challenge to Block's strict criteria for cognitive access from Peter Carruthers (2015b). In an updated version of his account of consciousness, Carruthers suggests cognitive access need not involve encoding in working memory at all. Instead, with clear echoes of Prinz (2012), he suggests that what is involved in both consciousness and cognitive access is the use of attention to boost sensory signals and cause them to be globally broadcast. Next, he claims that the degree of attentional amplification required to cause a given representation to be globally broadcast is significantly less than the degree of attention required for encoding in working memory. As he puts it "more attention [is] necessary to sustain a representation in working memory for purposes of reporting than is needed to result in the global broadcast of the corresponding perceptual representation." His final crucial claim is that global broadcast by itself is a perfectly adequate notion of cognitive access. Global broadcasting, in the absence of encoding in working memory, may serve to make representations only briefly available, but constitutes a genuine form of access.

Some parallels with my own account and that of Prinz are clear: both I and Carruthers take consciousness to consist in a process prior to encoding in working memory, but subsequent to localized sensory states. However, there are some major differences. First, unlike Carruthers, the account I have given makes no reference to attention. This reflects a careful choice on my part, since (as discussed in Appendix 1), the empirical literature on relation between attention and other forms of memory is still highly vexed. A second important difference is that while Carruthers elsewhere (2015b) allows for the possibility of conceptual binding of sensory states, he does not take it to be a necessary condition of a state's being conscious; sensory states alone can be globally broadcast.

Most importantly, however, I would suggest that the CSTM account is on a firmer footing than Carruthers' theory when it comes to the mechanisms of consciousness. In particular, Carruthers faces a serious challenge concerning his construal of the global workspace. He claims for example, that "it is possible to make sense of the idea that phenomenal consciousness is richer than we can report in a way that is consistent with the claim made by Dehaene and others that phenomenal consciousness coincides with global broadcasting." However, it is highly doubtful whether Dehaene himself would agree with this. As noted, he takes it to be the case that only a single item is globally broadcast at a given time. His motivations for this claim stem not from claims about working memory specifically, but from a wide range of neural and psychological data concerning the capacity limits of the workspace and the manner in which individual states are selected for broadcast. If Carruthers cannot rely on Dehaene's own specification of the global workspace, then it is not clear what sort of empirically well-grounded psychological process he can appeal to as being constitutive of rich phenomenal consciousness. By contrast, the Workspace-Plus theorist has a clear story to tell here in the form of high-capacity CSTM.

While there are important commonalities here, I remain doubtful of the idea that there is an empirically grounded way to extend the notion of cognitive access to include states that are not encoded in working memory or broadcast on Dehaene's Global Neuronal Workspace. This suggests to me in turn that the Workspace-Plus theory should be classed as an access-independent theory, although a somewhat novel one.

## 4.8 – Conclusion

In this chapter, I have had two main goals. The first has been to spell out some of the major theories concerning the mechanisms of consciousness, and to illustrate some respects in which they might be held to be unsatisfactory. My second goal has been to show how CSTM might usefully contribute to this debate, and to offer a new model of conscious experience in the form of the Workspace-Plus model. The outline of how this theory might function is only preliminary and open to further elaboration and revision. Moreover, it leaves many important questions about conscious experience unsettled, such as the nature of intentionality and the metaphysical status of consciousness. However, I hope that I provided reasons for taking the Workspace-Plus Model seriously, and for regarding it as a promising candidate theory of consciousness.

In the next and final chapter, I will focus on some unresolved issues concerning consciousness, CSTM, and its place in nature. In particular, I will consider how the Workspace-Plus model can be located within Burge's account of the architecture of the mind. I will go on to examine in more detail whether we have reason to think that non-human animals may possess an equivalent of CSTM, and whether we can apply the theory to one challenge in particular, namely the assessment of conscious suffering in non-human animals.

# CHAPTER 5: CSTM and its Place in Nature



## 5.1 – Introduction

Before heading into this final chapter, it is worth taking stock of the main claims I have made thus far. I began in Chapter 1 by summarizing the main existing frameworks for thinking about short-term memory and presented a synthesis of these theories in the form of the Sensory-Cognitive model. I then sketched how this framework might be relevant to debates in philosophy and cognitive science. Next, in Chapter 2, I examined a range of recent evidence to suggest that the Sensory Cognitive Model may be missing out an important stage of processing in the form of Conceptual Short-Term Memory. In Chapter 3, I explored the idea that CSTM may be the mechanism responsible for categorical perception, and building on this suggestion, I proposed ways in which it might inform debates about high-level phenomenology and cognitive penetration. Next, in Chapter 4, I examined the theories of consciousness debate in more depth, and showed how CSTM might be a key component in a Workspace-Plus theory of consciousness. Such a theory, I argued, could offer a number of independently attractive features as an account of conscious processing.

In this final chapter, I wish to turn from specific debates in philosophy of mind and cognitive science and attempt to locate my claims about CSTM within a broader set of philosophical and scientific concerns. As an important initial goal, I will spell out how the kind of account I have offered might be understood within the framework of Tyler Burge. This may make it easier to make sense of my proposals in a broader theoretical context and allow for easier comparisons to be drawn to other theories. Building on this discussion, I will also examine whether non-human animals may have some analogue of CSTM. I take this to be an especially important consideration given that I have already suggested that CSTM and Central Cognition might together constitute the psychological basis of conscious experience in human beings. If so, questions about animal consciousness may rest on the presence or absence of some faculty akin to CSTM. As a final goal for this chapter, I will examine one particular speculative but promising application of the theory, namely as a framework for assessing experiences of suffering in non-human animals.

I will proceed as follows. I begin in 5.2 by laying out Tyler Burge's model of perception and how it carves up the mind. I go on in 5.3 to locate CSTM within this framework, and suggest it is naturally considered a form of post-perceptual processing. In 5.4, I consider whether we have reason to think that non-human animals might have an equivalent faculty, and go on in 5.5 to consider whether this faculty may serve as the basis for consciousness in nature more broadly. Finally, in 5.6, I lay out a proposal for using my theory of consciousness to assess suffering in non-human animals.

## 5.2 – The Burgean framework

Questions about the existence of mentality and consciousness in animals have been a source of considerable philosophical dispute since the inception of philosophy (Sorabji, 1993). In the Early Modern period, a particularly contrast of philosophical opinion can be seen in a comparison of

Descartes' famous claim that animals were automata (Regan and Singer, 1989: 13-19) with Hume's insistence that "no truth appears to me more evident than that beasts are endowed with thought and reason as well as man" (Hume, 1968). In the twentieth century, skeptics of animals minds such as Davidson (1975) continued to deny that animals have mental states with intentional content, even as scientists were developing increasingly sophisticated models of animal cognition (Tolman, 1948). In the last few decades, however, it seems that a growing consensus has emerged among philosophers that we cannot simply remain silent about the minds of animals (Bermudez, 2003). Yet even among commentators who allow for animal thought, there is still considerable debate as to whether animals have any mental representations that might justifiably count as conceptual (see Chater & Heyes, 1994).

The very idea that animals might have *Conceptual* Short-Term Memory, then, is a matter of deep controversy. As I will spell out more clearly in this chapter, however, I think the accounts of CSTM and consciousness that I have given thus far can be understood in a way that allows for their broad applicability to a potentially wide range of non-human animals. This may seem surprising: the kind of experiments that I suggested provide our best evidence for CSTM involved learned high-level categories such as 'wedding', 'picnic', and so on, which most animals surely lack. However, we should distinguish the question of what counts as the *best evidence* for the existence of a buffer like CSTM from the question of what its *minimal* functional role may be. Thus, while I take it as constitutive of CSTM that it can encode information in respect of these high-level categories when they are available, it need not the case that such categories always *are* available (see the discussion in 4.6). This distinction may prove important in considering the possibility that CSTM or some close analogue exists in non-linguistic creatures (including quite simple ones) whose mechanisms of learning, especially socially-mediated and linguistic learning, are absent or highly limited. However, I recognize that for some theorists, namely those who are skeptical of attributions of sophisticated mental representations to animals, my arguments in this chapter may seem deeply unpersuasive.

With that caveat in mind, I now wish to consider how my proposals might feature in the important and influential framework of the mind's place in nature provided by Tyler Burge (2010b). Burge's account is one of the leading theories of perception, cognition, and content, and his account has the advantage of being closely involved both in scientific debates about animal minds and in theoretical debates about how we should carve up the mind. While I will not argue in defense of Burge's theory here, I will now lay out its basic commitments.

Consider a simple case of what we would naturally call perception: I notice a man pull up in an unfamiliar car outside my house. According to Burge, this kind of event can be broken down into a series of stages, each of which possesses its own distinctive properties (see Fig. 5a).[43] First, we can talk about the initial stage of sensory registration, in which low level visual mechanisms are activated by proximal stimuli. What distinguishes this stage of processing from perception proper is that, while such states invite explanation in terms of causal and statistical regularities with environmental stimuli, they can be (and, as a matter of scientific practice, Burge contends, typically are) fully explained without recourse to representational content. As Burge puts it, the "key feature of representation – that it can be accurate/inaccurate or true/false – plays no role in the so-called representation involved in information registration" (Burge, 2014b).

[43] Even this five-part account is somewhat simplifying, and is derived from the bare-bones version of Burge's view presented in Burge (2010a, 2014a) and Block (2014c). Burge's more developed account (2010b) includes numerous qualifications and elaborations that lie beyond the scope of this paper.

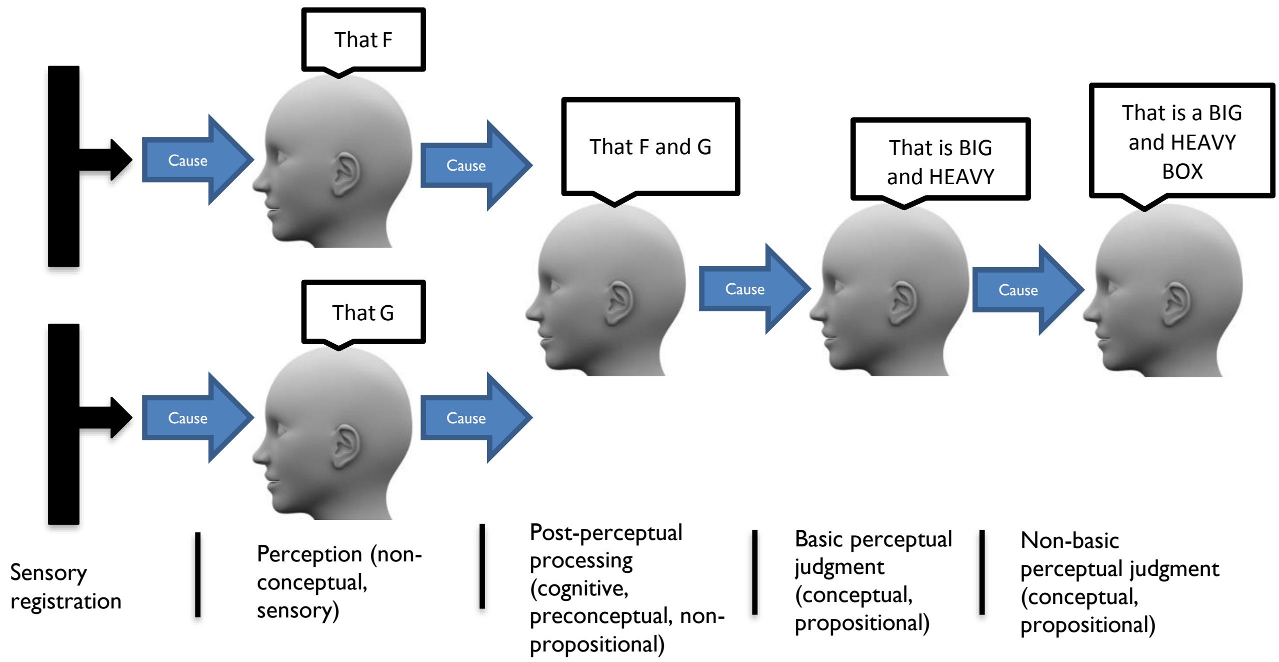



**Fig. 5a.** A schematic illustration of Burge's model of perception (adapted from Block, 2014c)

Second, we can distinguish genuinely perceptual forms of representation. Specifically, Burge claims, the step from information registration to full blown representation involves objectification; in Burge's terms, the "marking off [of] states that are as of specific system-independent elements in the environment from states idiosyncratic or local to the perceiver" (2010b). This is accomplished by *perceptual constancies*; that is, the ability of a perceptual system to maintain consistent representation of a given environmental property such as shape, location, lightness, or location.

Perception, on this view, has a distinctive syntactic structure and format. Specifically, Burge claims that perception constitutively involves a *singular referring element* and a *perceptual attributive*. The referring element provides the referent of the perceptual state, and "functions to pick out, identificationally single out, particulars in a direct, not purely attributive manner" (2010a). The perceptual attributive, by contrast, indicates some property or relation, such as a being a particular shape, a particular color, being in motion, or being a given shape.

Thus for Burge, and unlike, for example, Dretske (1981), perception constitutively involves perceiving things in some way or another; as he puts it, "[p]erceptual states must represent what is perceived as being a certain way" (2010a). Another crucial point for Burge is that he takes perception, in contrast to the full-blown propositional contents characteristic of thought, to be incapable of standing in inferential relations or of representing more abstract logical relations such as negations, quantifications, or conditionals.

Thus far, we can spell out what is involved in seeing the car in terms of processes of sensory registration carried out at the retina and in early vision, followed by the application of perceptual constancies to generate a bona fide perceptual representation of the car as an objects with a certain shape and color. The third key stage in a Burgean account of a mental event like noticing a car is a little more mysterious. Specifically, Burge claims that there are "a range of post-perceptual types of mental states and events (beyond perceptual anticipation and perceptual memory) that do not count as reason, but that figure in cognition... and represent at more abstract, intermodal levels" (2010a).

Burge himself takes such states to be "preconceptual" (Burge, 2014a), and as examples, he cites "[p]sychological maps for navigation and perhaps certain representations of magnitudes", as well as representations "formed through learning or other processing in long-term memory (modal or amodal)." Applying this again to the case of seeing the car outside, it might the case that these post-perceptual processes contribute to my representing of the car as being at a certain place in my local environment, for example.

It is also perhaps illuminating at this point to note briefly that Burge points to Susan Carey's notion of *core cognition* as an example of these states (Carey, 2009), though he does not endorse her specific theory without qualification. While a proper treatment of Carey's account lies far beyond the scope of this chapter, the central idea is that humans possess an fundamental system of concepts for thinking about the world that functions to represent various basic properties and relations; in Carey's terms, its domain is "middle-sized, middle-distant objects, including representations of causal and spatial relations among them", as well as agential and causal relations. Importantly, however, the classificatory activity performed by core cognition relies on innate categories: as Carey puts it, "the input analyzers that identify its referents are not the product of learning" (Carey, 2011). Carey also maintains that the contents of core cognition are both *iconic* and (contra Burge) *conceptual*.[44]

The final two stages of processing involved in an event like the one under consideration are what Burge terms *basic perceptual judgments* and *non-basic perceptual judgments*. Unlike perception or post-perceptual processing, judgments have a fully propositional structure, and can represent logically complex states of affairs such as negation. They are distinguished from one another insofar as the kind of attributions possible for basic perceptual judgments are limited to conceptual equivalents of the nonconceptual attributives already present in perception (though, as Block notes, unlike perceptions, they "need not be bound to a time and a place"; Block, 2014c). Non-basic perceptual judgments, by contrast, can attribute properties not present in perception, can arise independently of perceptual inputs, and can express various complex propositions (such as conditionals or disjunctions) that may not be possible in the case of basic perceptual judgments.

## 5.3 – CSTM and categorical perception in a Burgean framework

The above brief description hopefully served to paint a rough picture of Burge's model of the mind, and of perception in particular. One thing that I hope will be clear from the foregoing

[44] A number of theorists have suggested there is a tension between these claims, especially in the case of concepts such as *cause* or *agent*. See, Gauker, 2011, and Shea, 2011.

discussion is that it is far from obvious where we should locate categorical perception of the kind discussed in Chapter 3. I now wish to dwell on this question, and suggest a way in which a theorist sympathetic to Burge's picture might be able to integrate the CSTM model presented thus far.

Broadly speaking, there are three possible places in Burge's hierarchy of perceptual processing where categorical perception and CSTM could plausibly be located. One place is at the level of perception proper. This is a proposal developed by Block (2014a) and endorsed by Burge (2014a), and was discussed in Chapter 3. There are a number of reasons why I am not inclined to think that this proposal is fully satisfactory, however. I presented some of these doubts in Chapter 3, and will not repeat them here. Note again, however, that Block and Burge offer quite an austere picture about the kind of high-level attribution that can occur in perception. For example, Block states that "probably there are no culture-specific higher level attributives for teacups and recessions", while Burge insists that perception does not contain "attributives for cultural attributes like being a car."[45] To the extent that one is sympathetic to the idea that perceptual *experience* does indeed represent these high-level properties (as I argued in Chapter 3), then Block's proposal will not fully explain the phenomenon of categorical perception.

If we assume that at least some categorical perception is not perceptual, could it instead be located at the level of beliefs? I already gave some reasons in Chapter 3 for doubting whether categorical perception could be satisfactorily understood in terms of beliefs, at least insofar as we take beliefs to be central cognitive processes. One option might be to try to explain categorical perception as being a matter of Burge's basic perceptual beliefs. However, this is unlikely to help

[45] Block's two examples ("teacups and recessions") are arguably quite different. Teacups are concrete objects with a distinctive shape and distinctive social function, whereas recessions are abstract entities that involve a scattered class of objects and exist only over extended durations. It seems to me quite reasonable, as I suggest in Chapter 3, to claim that the former, but not the latter, can feature in the contents of perceptual experience.

us much: Burge himself takes basic perceptual beliefs to be highly limited in their content, “containing only conceptualizations of the perceptually basic attributives.” Given the constraints mentioned that Burge takes to apply to perceptual attributives, then, basic perceptual beliefs will not allow us to explain, for example, seeing a man as drunk, or an object as a teacup.

Could categorical perception then be a matter of *non-basic* perceptual belief? While I cannot not rule this out, there is a particular reason why the Burgean framework may make this proposal seem especially unattractive. First, note that all perceptual beliefs for Burge are full-blown propositional attitudes in good standing, and as such there are relatively demanding constraints that a creature must satisfy in order to count as possessing them. As Burge puts it, a “much more limited range of animals, including humans, have propositional attitudes, including propositional perceptual beliefs. There is evidence that apes have them. Probably several other non-human animals have them” (2010b). If Burge is correct in this assertion, and, further, we take categorical perception to consist in perceptual beliefs, then it follows that a similarly limited range of animals will exhibit categorical perception.

However, as will be discussed below, there is some tentative evidence that many fairly simple animals can engage in categorical perception. If we take this evidence at face value, and assume with Burge that propositional attitudes are relatively rare in nature, then categorical perception cannot be entirely a matter of perceptual beliefs. Some theorists, of course, may disagree with Burge’s insistence that propositional attitudes are possessed by a limited range of animals (see Carruthers, 2009), and I am open to this possibility. However, given Burge’s insistence that relatively few animals have propositional attitudes, it is unlikely that he would agree with them.

On a more general note, it seems to me that it would be a striking – and rather surprising – finding if it were to be demonstrated that all and only animals capable of holding propositional attitudes were capable of engaging in categorical perception. Abilities like reasoning, thinking, and inferring seem vastly more cognitively sophisticated than merely being able to perceive objects as belonging to a particular semantic category. There is arguably even an argument from parsimony in this respect: to the extent that we wish to follow Morgan's Canon and avoid interpreting "animal activity... in terms of higher psychological processes if it can be fairly interpreted in terms of processes which stand lower in the scale of psychological evolution and development" (Morgan, 1903), there is some reason to avoid interpreting categorical perception in terms of full-blown propositional attitudes if we have simpler psychological accounts available.

I conclude then, that there at least some reasons for the Burgean, at least, to look for a place for categorical perception outside of both perception proper and perceptual judgment. Thankfully, there is a natural place for categorical perception and CSTM within the Burgean framework, namely as a form of postperceptual processing. As noted by Burge, postperceptual representations include those "formed through learning or other processing in long-term memory" (2014a), and he argues that such representations may be highly abstract and non-domain specific. Moreover, by locating the representations involved in categorical perception at the stage of postperceptual processing, we can also allow for learning at the level of CSTM while accommodating the idea that the representations involved in perception proper are, as Burge wishes, largely innate.

We might incorporate CSTM within Burge's model, then, as a faculty for categorical perception that arises at the stage of postperceptual processing. However, one important problem for this proposal is Burge's insistence that postperceptual processes, as distinct from basic perceptual judgments, are non-conceptual. Burge's basis for this claim is that he takes concepts to be constituents of propositional attitudes like judgments, and not to occur outside of such attitudes.[46] Thus the claim that postperceptual attitudes possess only nonconceptual content is a direct theoretical consequence of the fact that he takes them not to possess a propositional structure.

I am highly sympathetic to claim that representations in CSTM do not have a fully propositional structure, consisting instead of, perhaps, a concept and a singular referring element (this is, in fact, the difference between $GW+_1$ and $GW+_2$). However, I have also insisted that CSTM is, true to its name, a mechanism for *conceptualizing* perceptual information. Does this scotch the idea that CSTM may be a form of postperceptual processing as defined by Burge?

I would suggest that there is perhaps less disagreement here than meets the eye. Burge can agree with the CSTM theorist that the relevant representations involved at this stage are characteristically abstract, high-level, and can be acquired through learning. Additionally, as already suggested, I take the best account of the contents of CSTM ($GW+_1$) to involve the idea that they are a kind of hybrid representation, involving both iconic and conceptual elements, and may not have a full-blown propositional structure, allowing Burge to maintain that there is a non-conceptual element operative in postperceptual processing.[47]

This may not be enough for Burge. Still, the nonconceptual version of the Workspace-Plus theory ($GW+_3$) might be fully acceptable within the strict terms of Burge's framework.

[46] See Burge (2010a). "Conceptual attributives, or predicates – as I use these terms – constitutively contribute to propositional contents... concepts–in my usage--are constant, freely repeatable elements in propositional representational contents."
[47] Cf. Burge's comment on Carey: "Carey calls these [core cognitive] systems components of core knowledge, and takes them to involve use of "concepts", she believes that these systems of representation are iconic, not propositional. So her terms 'knowledge' and 'concept' are terminologically different from mine."

Note, moreover, that there may be some scientific grounds for gaining traction in this dispute. In particular, it seems an open scientific question whether the same neural vehicles recruited for categorical representation in CSTM might also serve in different contexts as the vehicles for some of the contents of propositional thought. If that is right, then the *very same* neural vehicles that are recruited by CSTM to represent, say, 'that BLUE BOX' might also be able to be accessed by Central Cognition for the purposes of properly propositional thought ('that BLUE BOX would look good in my apartment'). If that is how the neural story unfolds, then Burge's insistence on the nonconceptual nature of postperceptual processing would rest solely on abstract philosophical rather than psychological considerations. Of course, it could also emerge that judgment, in Burge's sense, relies on representations of a fundamentally different kind from those that feature in postperceptual processing. This aspect of the debate, I would suggest, is best left to science to resolve.

## 5.4 – CSTM and Categorical Perception in Non-Human Animals?

The foregoing arguments have aimed to establish that, if we wish to locate categorical perception in the Burgean framework, it should be subsequent to perception proper but prior to cognition at the level of postperceptual processing. I have previously argued (in Chapter 3) that categorical perception in human beings is a function of CSTM, and (in Chapter 4) that CSTM may be one of two constitutive bases of consciousness.

Even granting these assumptions, however, it does not follow that categorical perception in animals is conscious. For one, it may be that the psychological mechanism (or mechanisms) responsible for categorical perception in animals are very different from CSTM. Recall also that I allowed for the possibility (in 4.6) that certain kinds of semantic processing (such as those involved in the early stages of linguistic understanding) may be performed by unconscious modular systems. It is eminently possible, then, that at least some apparent cases of categorical perception in non-human use similar mechanisms.

With these caveats in mind, I nonetheless wish to explore the idea that categorical perception in some animals is based in some relevantly similar mechanism to CSTM. If so, then by the lights of the proposals in the previous chapter, those animals would undergo conscious experience.[48] Note that in proceeding I will not attempt to argue for $GW+_1$ specifically, but rather the broader Workspace-Plus framework. For that reason, in much of following discussion I will use the more generic term "semantic" in preference to "conceptual", thereby remaining neutral on the merits of $GW+_1$ as opposed to $GW+_3$.

Given the many challenges involved in determining whether animals possess a given cognitive capacity, this will be largely an exploratory exercise: it would certainly be a quixotic endeavor to attempt to prove that animals have a form of memory identical to CSTM, not least because the existence of the latter in human beings itself is still a matter of controversy. Nonetheless, we can speculate as to whether animals may have some 'cognitive buffer' that could be broadly equivalent in function to CSTM in human beings. I take it that the minimal requirements for such a mechanism would be that it (i) was a postperceptual process that (ii) was integrated with an organism's perceptual systems, its memory, and any higher-level cognitive systems it possessed, (iii) functioned to encode incoming perceptual signals in a respect of semantic categories, and (iv) made its outputs available for the production of 'whole organism' behavior such as coordinated movement, mating, predator response, and gathering food (Burge 2010b: 326ff.).

[48] Of course, given the claim of the Workspace-Plus view that consciousness is associated both CSTM and with Working Memory, an alternative strategy for showing consciousness in animals would be look for evidence of the latter in animals. However, I will explore the former option here, in part because I take my claims about CSTM to be the most distinctive of my broader thesis, but also because I suspect that CSTM is a psychologically and phylogenetically more primitive faculty.

I take it that CSTM in human beings meets all these criteria. Of course, in the human case, the contents of CSTM include high-level learned representation of things such as picnics and weddings. In simpler animals, these semantic categories might be largely innately specified, attributing ecologically relevant properties such as ‘threat’, ‘food source’, and ‘mating partner’, as well as perhaps, environmental attributes such as ‘steep slope’, ‘cover’, and ‘open ground’, and so on.

This is just a simple sketch of the proposal. Discovering the specific categorization mechanisms available to individual species of non-human animals and determining whether these indeed rely on higher-level postperceptual processes or are a function of lower-level sensory or perceptual systems is a painstaking and extremely challenging endeavor, and will draw upon a wide range of disciplines. However, there is some evidence that categorical perception may be surprisingly widespread in nature. Of course, any such claims will be tentative, given that the need to exclude debunking hypotheses makes it notoriously difficult to make firm conclusions about animals’ cognitive capacities based on behavior alone.

However, some cases are highly suggestive. For example, note that pigeons can be trained to discriminate novel pictures based on whether or not they contain images of people (Herrnstein & Loveland, 1964), and also seemingly learn to discriminate novel beach scenes from scenes of mountains or streets (Kirkpatrick et al., 2014). Rats, too, can learn to make same-different discriminations on novel stimuli (Nakagawa, 1993), and can learn to distinguish novel pictures of chairs, flowers, cars, and humans on seemingly categorical grounds (Brooks et al. 2013). Cichlid fish who observe fights between pairs of conspecifics are able to recognize which fish are victorious in different bouts, and can subsequently keep track of those animals that have been more successful overall in combats with rivals, seemingly taking into account not merely brute win/loss ratios but more abstract hierarchical properties (Grosenick, Clement, & Fernald, 2007). Perhaps most strikingly of all, bees can be trained to make same/different and above/below distinctions for wholly novel stimuli (see Fig.5b, below, and Chittka & Jensen, 2011 for a review).

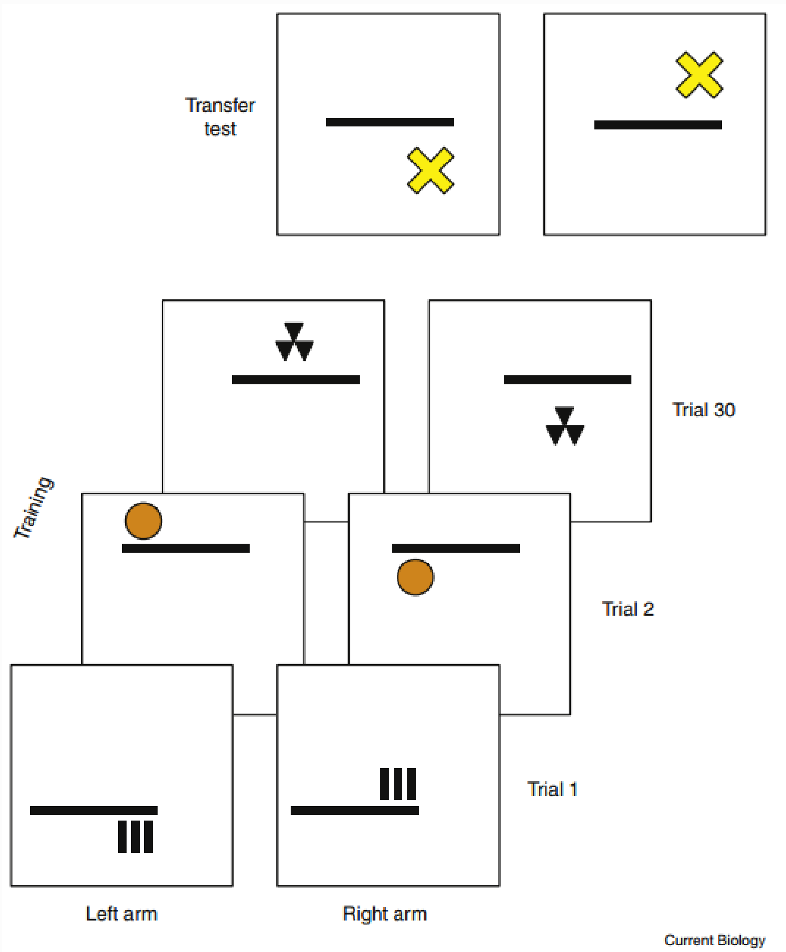


**Fig. 5b.** Bees can learn to navigate tunnels to find sucrose solution using novel cues. They are given a choice of two arms in a maze, and are trained to choose the one in which the stimulus occurred above the line rather than below it (another group was trained to choose targets below the line, and performed similarly). Stimuli and their positions were varied across trials to ensure that bees were not relying on fixed configurations of stimulus items. This learned ability then transferred to wholly new stimulus displays. Image adapted from Chittka & Jensen, 2011, discussing the work of Avarguès-Weber, Dyer, & Giurfa, 2011.

One might worry that some of the examples above just show that creatures can be taught to make certain kinds of visual discriminations; it does not show that they in anyway grasp the *significance* of the things discriminated. Thus Allen & Hauser (1996: 51) observe that "[i]t is possible to teach a human being to sort distributors from other parts of car engines based on a family resemblance between shapes of distributors. But this ability would not be enough for us to want to say that the person has the concept of a distributor."

While this kind of worry may not apply to all of the examples given (Cichlids' visual classification of rivals into winners and losers, for example, seems intimately integrated into a fairly complex range of social behavior), nonetheless the point is well taken. Certainly, a subject's ability to visually discriminate Xs from Ys does not show that they possess the *same* concepts of Xs and Ys that might be possessed by someone who has a broader understanding of the significance of the two categories. I would certainly not claim, for example, that pigeons share the *human* concept STREET based just on their ability to distinguish street scenes. Still, it certainly seems an open possibility that we may be able to attribute some *simpler* concept to pigeons in the form of a learned high-level attributive applied in perception to distinguish two kinds of visual scene.[49]

In any case, however, I am not for the time especially concerned to argue the case for concepts in non-human animals, but rather the possibility of categorical perception in the broader

[49] In this regard, it is worth noting that the use of the term 'concepts' to describe such categorization behavior is ubiquitous in the relevant scientific literature, occurring in the title of almost every paper cited earlier, although it is of course possible that experimentalists are using the term excessively broadly.

sense described earlier. In that regard, the capacity to reliably discriminate novel images from two different categories seems to involve *some* capacity for abstraction and higher-level categorization that I would claim reasonably warrants attributions of categorical perception, even in the absence of particularly sophisticated inferential abilities. I therefore take these cases to present at least good prima facie evidence for thinking that some non-human organisms engage in the classification of perceptual information in respect of higher-level features that cannot be spelled out merely in terms of the kind of largely innate categories provided by perceptual constancies. If so, then this would provide tentative for evidence for a mechanism that might satisfy criteria (i)-(iii) above. The fact that animals can use this categorical information to, for example, select one of two tunnels (as in Avarguès-Weber, Dyer, & Giurfa, 2011) shows that it is available for use in whole organism behavior, satisfying criteria (iv).

Of course, behavioral considerations can only be suggestive, and do not allow us to rule decisively about the kind of psychological mechanisms that may be involved. It may be possible, for example, that the psychological processing that underlies behaviors like the above is, contrary to first appearances, performed by some kind of fairly simple mechanism that is not well integrated into a creature's broader perceptual and cognitive systems, thereby failing to satisfy criterion (ii) or criterion (iv).

I certainly do not rule out this possibility. However, before leaping to this sort of debunking explanation, it is worth stressing that the categorization mechanisms used by even simple animals frequently exhibit a greater degree of sophistication and integration than might be initially supposed. Consider, for example, the alarm pheromones produced by ants to alert and recruit others to repel a threat (Hölldobler & Wilson, 1990). We might hastily assume that this sort of response would operate at a very simple level, and be a poor candidate for *bona fide* categorical perception in non-human animals. However, a closer look at the neuroanatomy of ants' response to pheromonal alarm signal reveals a much more complex picture. Neural activations caused by the detection of pheromonal alarm signals trigger distinctive patterns of firings that extend from sensory areas all the way to the ant protocerebrum, a multisensory area involved in integration, learning, and memory (Mizunami, Yamagata, & Nishino, 2010). The behavioral effects of such pheromones are also nuanced: the pheromones themselves do not themselves activate aggressive behavior, but instead serve to *sensitize* the ant to subsequent "non-pheromonal [visual] sensory stimuli associated with a potential enemy." The picture that emerges, then, is of a complex multi-sensory process that results in subtle shifts in ant's behavioral dispositions. In light of this complex neural and behavioral data, we should not be too quick to assume that even the lowly ant lacks an integrated faculty for categorical perception like the one described above.

Finally, while it would be a mistake to put too much weight on evolutionary considerations, it is easy to imagine how the ability to integrate sensory information from a variety of modalities into a unified format for the guidance of behavior might be of evolutionary value, even for creatures that lack propositional attitudes and possess relatively constrained and inflexible repertoire of behaviors. Such an ability might allow organisms to rapidly identify signatures of predators or of prey, for example, without relying on perhaps slower and more dangerous processes of associative learning. Interestingly, similar evolutionary considerations for the structure of categorical perception in human beings have been advanced by psychologists in explaining the category-specific nature of associative agnosia in human beings. Thus Caramazza and Shelton (1998) suggest that "evolutionary pressures have resulted in specialized mechanisms for perceptually and conceptually distinguishing animate and inanimate kinds (R. Gelman, 1990; Premack, 1990), leading to a categorical organization of this knowledge in the brain." I take it as a reasonable hypothesis, then, that many non-human animals may have some system like CSTM for integrating and categorizing information.

## 5.5 – Consciousness as Post-Perceptual Processing

Let us entertain the hypothesis, then, that many non-human animals possess CSTM or some equivalent semantic buffer. I have already mentioned some considerations in the previous chapter in relation to the human case that may make it attractive to identify this kind of processing with perceptual experience. However, even someone sympathetic to these considerations might doubt whether this kind of processing would be necessary or sufficient for consciousness in much simpler creatures. To that end, I now wish to examine whether we can offer some related reasons for thinking that consciousness more broadly in nature may be tied to postperceptual processing of the kind I take to be performed by CSTM.

An initial way to motivate this idea may be to examine the contrast between postperceptual processing and perception proper. To begin with, note that perception in its strict Burgean sense is not a good candidate mechanism for conscious experience simply because it can be accomplished by very simple mechanisms at the periphery of the architecture of the mind. Thus, as Burge notes, "there is evidence that some color constancies – hence perceptions – in bumblebees occur at the retinal level, with nearly no processing. Such constancies almost surely occur before consciousness could occur, even if the bees *are* conscious" (2010a).[50]

Burge's insistence that perception at the retinal level is not conscious seems highly

[50] There are nuances to Burge's account that I have not spelled out here. For example, he argues that while individual perceptual mechanisms may operate at the subindividual level, they are only rightly counted as perceptions insofar as they *contribute* in some way to individual-level perception (Burge, 2010b:369)..

persuasive. But why should we find it so immediately plausible to deny that processing at the retinal level should be conscious? I would suggest two main reasons. The first is that our notion of conscious experience is closely bound up with existence of a single unified perspective or point of view on the world (Nagel, 1974), and it is hard to see how peripheral states like retinal activations could directly contribute to such a perspective. While retinal processing might *causally* contribute in some way to a bee's awareness of the world, this would only in virtue of further downstream processes that allowed its outputs to combined and integrated into a coherent set of conscious representation that could jointly form a bee's experience of the world. This suggests to me that if we are looking for a place to locate conscious experience, it should be at a level of processing in which information from different domains could be combined into an integrated whole.

A second closely related reason that constancy mechanisms at the retinal level may not seem like plausible candidates for consciousness is that their content is not in a form that can directly guide an organism's goals and behavior. In the previous chapter, I argued that a fundamental connection between consciousness and categorical perception in the human case could explain the way in which that our perceptual experiences are intrinsically meaningful for us, carrying information not just about the sensory properties of objects but their relevance to our interests (cf. Gibson, 1977). Indeed, as suggested by the near synonymy of 'consciousness' and 'awareness', there is some attraction to the view that conscious experiences fundamentally serve, in some way or another, to inform about us the world, our body, or the contents of our own minds, even if the resulting awareness is fleeting or indistinct.[51]

---

[51] I take this kind of consideration to be a large part of the appeal of the Intentionalist program in the philosophy of perception. Of course, most Intentionalists are committed to far stronger theses about the

Applying this very crudely to the bee case, then, I would suggest that it does not directly matter to the bee whether a given color gradient on a flower is a genuine color boundary or the effect of shadow; what matters is how this contributes to a representation of the kind of flower it is, and, for example, whether it is one that is a good source of nectar. While detecting color boundaries will of course be part of this, the deliverances of perception will only form a meaningful part of a bee's awareness of the world once they are located within a broader set of categories that are relevant to it behaviors. Putting this more generally, it seems a plausible hypothesis that consciousness arises only once perceptual information is passed on to mechanisms that can locate it within a broad set of ecologically relevant categories that are tied to whole organism behavior. Some of these may be immediately motivating, such as the recognition of an organisms as predator or prey. Other categories may serve to inform goals less directly; recognizing something as a landmark, a conspecific, or as an environment that has previously been rich in food, for example.

With both these considerations in mind, I would suggest that if animals possess some close analogue of CSTM, this might be a reasonable place to locate perceptual experience. Consider first the idea that consciousness involves a unified perspective on the world. I have suggested one of the defining features of the kind of semantic buffer described earlier would be that its outputs are partly encoded in respect of some unifying format (whether this is strictly conceptual or some form of intermodal nonconceptual content). If this is the case, then we can make sense of how such representations could form a coherent phenomenal whole. Similarly, the consideration that conscious experience consists in ecologically meaningful representations is a natural consequence of the kind of semantic buffer view described. The function of such a buffer would precisely be to encode sensory information in

relation between phenomenology and content than any I have endorsed here.

respect of the kind of categories that are relevant to the selection of behavior at the whole organism level. The point of categorizing a given animal as a *predator* or *potential mate*, for example, is to cause or motivate a creature to engage in a broadly appropriate behavioral response.[52]

The kind of animal experiments described in the previous section seem at least prima facie to meet these criteria. When a Cichlid fish avoids one rival (but not another) on the basis of having previously seen it dispatch numerous foes, it is presumably doing so because it has in some way or another previously categorized its rival as having successfully defeated other fish, and is subsequently able to use these categories to avoid a bruising encounter. When a rat or a pigeon comes to be able to discriminate novel scenes, it similarly seems to have learned that certain categories of picture reliably predict food rewards. Even the honeybees described by Chittka and Jensen, in their ability to accurately select one tunnel rather than another on the basis of novel arrangements of shapes, seem to have acquired some system of visual categorization that allows them to recognize that one tunnel will result in a better outcome.

Of course, as noted earlier, this kind of categorization may somehow turn out to be mediated by a peripheral mechanism that does not involve integration of information into a unified format. But if it is indeed subserved by some more integrated central mechanism akin to CSTM, I think we should take seriously the idea that these behaviors may point to conscious experience. The idea that such simple creatures as honeybees could be conscious may to some readers seem outlandish. There are certainly many other theories of consciousness such as

[52] There are also undoubtedly mechanisms that bear a superficial resemblance to semantic processing but are wholly explicable solely in terms of pre-perceptual sensory mechanisms. Consider, for example, taxis away from sodium chloride concentrations by paramecia; in such cases, there is no reason to suggest that the organism represents the sodium chloride *as* anything at all. For a thoughtful discussion, see Burge 2010b: 315-9.

higher-order thought views that offer a drastically different assessment of the distribution of consciousness in animals, and I have said nothing in this dissertation that comes close to decisively refuting the views.

However, as a final consideration, is worth briefly noting that my account bears some important commonalities to an emerging family of approaches to animal consciousness advanced by Bjorn Merker (2005) and Barron & Klein (2016). The approaches take their primary inspiration from comparative neurobiology, and are similarly expansive in their view of the range of non-human animals that may be conscious, extending the franchise of subjective experience to include a range of invertebrates including, in the case of Barron and Klein, honeybees and other insects. These theories also have important theoretical parallels to my own account, insofar as they take conscious experience to be based in systems intermediate between perception and high-level cognition, arising via the translation of perceptual and sensory information into a single “convenient format”, allowing for “multisystem coordination... ranking of behavioral priorities, and decision-making” (Merker, 2005).

There are, however, important differences between these accounts and my own. Both, for example, stress the representation of body as key to consciousness, and do not make any specific commitments regarding the format that is shared among conscious experiences. However, I think it is nonetheless helpful in evaluating the kind of account I have offered to observe that it can loosely grouped together with these other approaches into a broader family of theories, all of which take conscious experience to be neither a strictly cognitive nor perceptual processes, arising instead via the postperceptual integration of information into a single format.

## 5.6 – Applying the theory: CSTM and animal suffering

Thus far in this chapter, I have attempted to do two things. The first was to locate categorical perception and by extension CSTM within the Burgean picture of the mind, suggesting that it is a form of postperceptual processing. The second was to suggest that the CSTM theory of consciousness *qua* postperceptual account offers us a potentially promising way of understanding the place of consciousness in nature, and has important commonalities with other important contemporary approaches.

My final goal for this chapter will be examine in more depth how the picture of consciousness presented thus far might apply to one particular form of experience, namely suffering. Although I will initially consider the problem in its most general form, I am particularly concerned to give an account of suffering that might allow us to better gauge the presence or absence of unpleasant experiences in non-human animals.

While this discussion will take me away from the other concerns of this chapter, it is my hope that it will demonstrate the utility of my account for more practical problems in philosophy and cognitive science. The specific proposals I make provide a natural fit for the particular view of consciousness presented thus far, and hence may lend further credibility to some of its more tentative claims, as well as providing a case study in how my account might grapple with a particularly problematic form of animal experience.

Before developing my account, however, it would be helpful to clearly state what I will be referring to in speaking of suffering. In short, I will count a psychological state as constituting a case of suffering just in case it *feels bad*. The notion of 'feeling bad' is, I take it, intuitively clear: pain, nausea, severe hunger and thirst, and dyspnea (the inability to breathe) all typically feel bad to some degree, as do a range of different affective states such as anxiety, dread, and stress.[53] A theory of suffering, then, should aim to tell us what it is that these states have in common at the psychological or phenomenological level, as well as suggesting neurophysiological or behavioral correlates of suffering.

Among the various kinds of mental states, suffering, I would also suggest, is an especially important target for theoretical investigation. For one, it is directly connected to ethical and pragmatic issues: I take it that, *ceteris paribus,* undergoing states that feel bad is detrimental (to at least some degree) to an individual's welfare. And while we can often rely on reports about suffering to inform our ethical behavior towards other humans, this is not always the case: infants, patients in persistent vegetative states, and individuals under powerful anesthetics might well undergo experiences of suffering while being unable to tell us about it. The case of non-human animals is, however, perhaps the clearest application for a theory of suffering: not only can we not rely on report in most cases, but neuroanatomical and behavioral differences make it harder for us to rely on indirect measures to assess suffering. This gives us clear practical and ethical reasons for attempting to give a theory of suffering. [54]

How should we go about developing a theory of suffering? I would suggest that there are two main sources of information that can guide our theorizing. First, we can make use of pretheoretical and phenomenological knowledge about suffering, and second, we can look to

[53] There are all sorts of further complications here. For one, some ethical theories will allow that an individual can suffer in ways that don't impact their psychological states in any way, perhaps through our desires being unfulfilled, or through being deceived or misinformed (Carruthers, 2005; cf. Nozick, 1974). Another concern comes from Dennett (1996), who suggests states that feel only *slightly* bad (like a stubbed toe, or a mild hunger pang) do not constitute cases of suffering at all. I will not go into these worries further here. Instead, both objections can be set aside at least for the time on the grounds that it is plausibly at least a *necessary* condition for an individual to have an *experience* of suffering that they undergo an experience that in some way or another feels bad, thus we have some independent interest in providing a theory of such states.

[54] In light of this, it is striking that relatively little philosophical attention has been given to developing a theory of suffering. Some notable exceptions are Parfit (1984), Korsgaard (1996), and Kahane (2005). However, these theorists all approach suffering from a strictly theoretical angle, and do not closely engage with empirical work. More empirically-focused enquiries into the phenomenon of unpleasantness are undertaken by Shriver (2014) and Corns (2014) both in relation to felt unpleasantness of pain specifically. Some suggestions for understanding unpleasantness as a general phenomenon are given by Klein (2015), though he stresses that they are tentative.

scientific research, particularly research on animal behavior. I will now review briefly these in turn.

Consider first what we know, pretheoretically and phenomenologically, about suffering. I would suggest, in this regard, that four features of suffering experiences particularly stand out. To begin with, suffering experiences can be associated (as noted earlier) with a *wide range* of different states, ranging from clearly bodily sensations such as pain, nausea, and hunger to more complex affective states such as panic, dread, and disgust. While each of these states feels different, they possess something in common, namely that they are feel bad. Any theory of suffering we give should be able to elaborate on precisely what this shared 'badness' amounts to.

A second consideration is that suffering experiences do not seem to occur in isolation; it never happens in normal life that we have an unpleasant experience that entirely lacks any sensory or affective component. Rather, whenever we feel bad, there is *some particular state* that feels bad.[55] Note that this is not simply a matter of unpleasantness always occurring in association with a given sensation. To illustrate this, simply pinch your arm to the point that it starts to hurt. The feeling of badness, I would suggest, is not some further experience you undergo additional to the pain, but is intimately bound up with the feeling of pain itself. In other words, when a sensation feels unpleasant, the locus of the feeling of unpleasantness is just the sensation itself. Again, a theory of suffering should be able to explain why this is.

A third feature of experiences of suffering is that they are, broadly speaking, *commensurable*. If you are asked whether you would rather have a three week-long headache or a ten minute spell of nausea, you will quickly judge that the latter will be less unpleasant. Of course, some cases are harder to assess (a week-long headache versus a day of intense nausea,

[55] Again, there are some difficult cases here. For example, Grahek (2008) discusses a case where a patient reports an unpleasant feeling somewhere in his arm, but was unable to specify whether it was a pain or some other sensation. Similarly, one might wonder about how to characterize extended periods of negative well-being such as depression.

for example), but even when it is very difficult to gauge which of two experiences would be worse, we can still at least try to project ourselves imaginatively into the two experiences to assess which would be more unpleasant.

A final feature of suffering that is perhaps more controversial is its connection to motivation. I would suggest that unpleasant experiences are constitutively *motivating*: we want them to stop, others things being equal. Of course, this does not mean that we never voluntarily engage in experiences that feel unpleasant: someone on hunger strike may feel horribly unpleasant hunger pangs but refrain from eating because of an overriding desire to engage in a political protest, while a masochist may find pleasures in the experience of pain that more than compensate for the unpleasantness involved. Nonetheless, it is hard to even make sense of the idea that someone might deliberately seek out truly unpleasant experiences just for their own sake (rather, than, say, for novelty, or for a desire for self-punishment).

These are some considerations that a theory of suffering should attempt to accommodate. However, as noted above, there are also scientific considerations that should constrain our theory. Consider, for example, the phenomenon of pain asymbolia (Grahek, 2008). Patients with this condition undergo pains yet do not seem to find them unpleasant, despite being able to speak quite eloquently about the character of pain sensations they are experiencing. This might tell against a theory of suffering that took unpleasantness to be a basic feature of conscious nociception, for example.

A further kind of scientific constraint comes from the wide range of responses to deleterious stimuli that can be accomplished seemingly without suffering being involved. A simple theory of suffering, for example, might suggest that aversive behavior might provide good evidence of suffering. However, plants and bacteria engaged in quite complex aversive responses to damaging or dangerous stimuli, but presumably do not suffer (see Karban, 2008, and Lyon, 2015). Another sort of account might point to behaviors such as guarding injured limbs or licking of wounds as strong evidence of suffering, but there is data that suggests such actions are merely “reflexive responses mediated by brainstem”, on the grounds that they can be performed by decerebrate rats (Cha et al. 2012) and chickens (Gentle et al., 1997).[56]

There is even evidence that some relatively sophisticated psychological processes such as instrumental learning can sometimes occur without conscious sensation. For example, as briefly noted in Chapter 4, rats with transsected spinal cords are still capable of instrumental learning just via the isolated section of spinal tissue. Specifically, they can be trained to keep their rear legs elevated to avoid electric shocks, even though their rear legs cannot communicate with the rat’s brain at all (see Allen et al., 2009 for a review). Thus a view that simply identifies suffering with negative reinforcement must seemingly say that rats’ spinal cords can suffer all by themselves – surely an intolerable idea.

In summary, then, a theory of suffering should accommodate, as far as possible, the phenomenological and intuitive constraints mentioned above, and should be framed in terms that allow it explain why behaviors like those mentioned above that are not plausible candidates for suffering. Are there any approaches that can satisfactorily do this?

Developing the account of conscious experience given earlier in the chapter, I would suggest that unpleasantness specifically arises at the level of postperceptual processing, and occurs when sensory or affective states are conceptualized in respect of some simple amodal

[56] Of course, just because something is a reflex response does not mean that it does not also give rise to an experience of suffering. However, I take it that these examples provide at least *some* evidence against the idea that pain-guarding behavior is a reliable signature of suffering.

category with the content BAD-TO-SOME-DEGREE.[57] This may seem a circular proposal, relying as it does on the concept BAD. However, this is just a simple label for a concept whose content can be elaborated in a way that does not rely on a prior notion of unpleasant. More specifically, I think that the content of the concept can be cashed out in *motivational* terms as a state that is *to be avoided to some degree*. Some care is in order here, however. I do not want suggest, for example, that a rat is aware of its sensations *as such*. Rather, I would suggest that, following the general schema of $GW+_1$, the content consists of a singular referring element and a negatively motivating conceptual component; schematically, (That) *to-be-avoided-to-some-degree*.

To see how this might work in a given case, imagine that a rat undergoes some nociceptive sensory state. That nociceptive signal, I would suggest, then acquires subjective negative valence – felt unpleasantness – once it passes to postperceptual mechanisms that serve both to encode the sensation in respect of the concept *to-be-avoided-to-some-degree* and cause it enter conscious experience.

Though speculative, this account has a number of attractions. Most important, I would suggest, is that it can be cashed out in terms of a specific experimental paradigm that offers an independently intuitive test of suffering, as I will say below. However, it also handily accommodates the intuitive and phenomenological considerations given earlier.

Consider first the fact that suffering is a feature of many different types of experiences. This aspect of suffering experience would be explained by the fact that the same amodal concept

[57] If readers balk at the idea that rats might possess concepts, then the foregoing can be recast in terms of nonconceptual encoding. As noted in 4.4 and 5.3, I incline towards a very liberal notion of concepts, but for present purposes, at least, I do not take too much to depend on the issue.

BAD could be bound to a variety of bodily sensations and emotions.[58] Similarly, this account could explain why suffering always seems to involve some particular sensation or emotion that is the locus of suffering. The concept BAD, on this view, would occur only in conjunction with some further sensory or affective state to which it was bound.

This account would also explain the motivational nature of unpleasantness, as well its commensurability. On this account, to feel something as unpleasant is to conceptualize a sensory state as one that is *to be avoided to some degree*.[59] Thus for a subject to feel a state as bad would constitutively involve its being represented as a state that they should try to diminish or terminate. Our ability to compare such states in respect of their unpleasantness would thus in turn rely on our ability to become introspectively aware of how that state was conceptualized, specifically its degree of to-be-avoidedness.[60]

Though I have just provided an outline of an account, I take it that it provides a good initial fit for many of the considerations advanced earlier. Additionally, however, it provides a natural way of interpreting some crucial empirical data relating to *motivational tradeoff* behavior in non-human animals. The basic phenomenon of motivational tradeoff concerns the tendency of

[58] Could the concept BAD also be extended to states besides sensations and emotions, for example, perceptual states? This seems an open empirical possibility. On the one hand, it may be that the mechanisms for encoding in respect of BAD are hard-wired to only take certain kinds of sensory states as inputs. However, it may also be possible to subjects to learn to associate the concept with previously neutral stimuli. Consider the notorious 'Little Albert' experiments conducted by Watson (1930), in which an infant was caused to experience strong aversion to rats. It may be that through association, the experience of seeing a rat came to be directly unpleasant for Little Albert. Alternatively, it's possible that any felt unpleasantness in this case came from his undergoing unpleasant *emotional* responses to rats.

[59] As noted in the previous chapter, I think it likely that the content of many states in CSTM may be cashed out in imperatival terms such as these. See Martinez, 2011, Siegel, 2014, and Klein, 2015.

[60] One complication worth mentioning here is how to make sense of the notion of "to some degree" in conceptual terms. On the face of it, the reference to degrees might seem to suggest an analogue and hence nonconceptual format. However, it also seems a viable hypothesis that the relevant degrees of badness could be cashed out in conceptual terms. Very crudely, the relevant semantic categories might be "bad-to-degree-1" "bad-to-degree-2", etc.. If this seems an unlikely proposal, note that there are cases in which two unpleasant stimuli can be selected such that an animal is strictly ambivalent between them (Dawkins 2012: 150-75), choosing at random. This would not be an obvious prediction of a sensory account that cashed out unpleasantness in strictly analogue terms, since one unpleasant sensation would presumably always feel slightly worse than another.

some organisms to willingly undergo one negative state in order to avoid another even more negative one (or to achieve a desired goal) or come to highly value a previously neutral stimulus if it will relieve a negative stimulus. This provides us with a way of assessing an animal's preferences, and to gauge which sensations or emotions animals most strenuously wish to avoid (see Dawkins, 2012, for a review of animal preferences).

Motivational tradeoff behavior in various forms has been demonstrated in a wide range of animals across multiple phyla. For example, rats, in a reversal of normal preferences, will prefer a light chamber to a dark chamber in order to avoid unpleasant mechanical stimulation of an injured paw (LaBuda & Fuchs, 2000). Broiler chickens are normally strongly disinclined to jump over high barriers unless food-deprived, but will do so quite spontaneously without any food deprivation in order to get away from highly-crowded enclosures (Buijs, Keeling, & Tuyttens, 2011). Zebrafish normally prefer to swim to an environmentally enriched chamber rather than a barren, brightly lit one, but their preferences are reversed if the fish are injected with an irritating acid and the bright chamber is filled with an analgesic (Sneddon, 2011). Likewise, hermit crabs will more slowly vacate a shell following electric shocks if the species of shell is highly preferred (Elwood & Appel, 2009).

Speaking for myself, I find it hard to avoid interpreting cases such as these as providing some intuitive evidence for the presence of suffering. The capacity of organisms not merely to respond to bodily damage or other negative states, but to tailor their behavior flexibly in accordance with the relative intensity of such states bespeaks a cognitive sophistication that makes attributions of genuine felt unpleasantness seem far from scientifically absurd. Indeed, the notion of motivational tradeoff has previously been suggested as a scientific criterion for assessing conscious pain in non-humans animals (Sneddon, et al., 2014). Moreover, I know of no data suggesting that very simple creatures (that are unlikely candidates for conscious experience in the first place) can engage in motivational tradeoff.

Importantly for present purposes, the accounts of consciousness and suffering provided earlier would naturally suggest motivational tradeoff as a likely behavioral signature for suffering. Any organism that can encode its sensations in respect of how strongly they are to be avoided can presumably tailor its behavior accordingly to select the least-bad such state (indeed, it is hard to see what function this encoding states would serve if not to enable such behavior). This ability need not require any sophisticated inferential abilities: a creature may implement a simple behavioral rule that it causes it to attempt to avoid or diminish whichever state is encoded as more negative. By the same token, the account provided earlier might independently claim that the *inability* of a creature to flexibly select among different negative states suggested that those states were not being represented in a format that allowed for ready comparison between them.[61]

If the above proposals are correct, does it follow that crabs can suffer? While some may scoff at the very idea of suffering in relatively simple creatures such as crustaceans, I take it to be an open scientific possibility.[62] It may emerge, for example, that the apparent motivational tradeoff behavior observed in hermit crabs is performed by a quite different mechanism from that used by rats and pigeons, and that their capability to select among different negative states may be correspondingly limited. This might in turn suggest that we should draw furthers distinctions among different types of motivational tradeoff. However, I would suggest that, in light of the

[61] Nonetheless, it remains possible that there are yet more primitive forms of suffering. A creature, may for example, be able to encode at a fairly centralized level certain states as to-be-avoided, but be unable to encode as to-be-avoided to any particular degree. I leave this as an open question.

[62] I also do not take it to be particularly counterintuitive to suggest that crustaceans may suffer. Note, for example, David Foster Wallace's famous and popular essay, *Consider The Lobster*: "[W]atching the fresh-caught lobsters pile over one another… it is difficult not to sense that they're unhappy, or frightened, even if it's some rudimentary version of these feelings" (2005).

theory I have given, the experimental paradigms described above offer a scientifically tractable approach for further understanding of suffering.

## 5.7 – Conclusion

In this final chapter, I have endeavored to locate the proposals advanced in previous chapters within a set of frameworks and methods for understanding the mind's place in nature. I began by outlining the account of perception and cognition developed by Tyler Burge, and argued that within this framework CSTM is best considered a form of post-perceptual processing. I then suggested that the strategy of locating conscious experience at this particular point in the mind possessed independent attractions, and had much in common with other emerging approaches to the study of consciousness. Finally, I examined how my theory might be applied to the specific question of suffering experience. I argued that it could be used to give an independently attractive theory of suffering and argued that it was well placed to accommodate the rich body of data relating to motivational tradeoff behavior in particular. In summary, then, I hope that the proposals of this chapter have shown how the broader ideas of this dissertation can be embedded within the leading frameworks of the architecture of the mind and its place in nature, and contribute to important ongoing scientific and philosophical debates concerning experience in non-human animals.

# CONCLUSION

Thus concludes this dissertation. To summarize my main goals, I hope I have clearly presented the current evidence that may point to a new form of short-term memory, CSTM. I have also explored how this form of memory may shed new light on existing debates about categorical perception and consciousness. Along the way, I have also been able to explore how CSTM might also inform a number of other interesting and important philosophical discussions, including those relating to high-level phenomenology, cognitive penetration, the reportability of conscious states, and the experience of suffering.

One important feature of this dissertation is that its main ideas are framed around a still controversial empirical possibility, namely the existence of CSTM. I say now in closing that I take there be considerable philosophical value in exploring the significance of live empirical possibilities, even if the data ultimately rules against them: given the rapidly developing nature of scientific psychology, philosophers who waited who till all the science was settled would, I suspect, find themselves waiting a long time. Moreover, while the CSTM hypothesis has provided the backbone and central framework for this dissertation, I hope that many of my reflections and arguments would stand in their own right. For example, I claimed that categorical perception relies on some faculty intermediate between perception and Central Cognition, and while I appealed to CSTM in this regard, it is possible some other mechanism may do the trick. Likewise, the core framework of the Workspace Plus model and the suggestion that consciousness is a postperceptual process may similarly survive empirical data that cast CSTM into doubt.

Still, it is my firm hope and, dare I say, expectation that evidence for CSTM or some very similar buffer will continue to mount. But while the fate of CSTM and some of my specific claims ultimately lies beyond the frontier of philosophy, I hope that many of the proposals I have given in this dissertation will be of independent interest and value.

# APPENDICES

## Appendix 1: Attention & the Sensory-Cognitive Model

What is the relationship between Central Cognition and Attention? Numerous theorists, including Baars (2003), Baddeley (1992), and Cowan (2001) have taken attentional control to be a function of Central Cognition, and to be limited by its sparse resources. This is a major theoretical claim, and properly evaluating it requires that we give some account of attention itself. Notwithstanding William James's (1890) claim that "[e]veryone knows what attention is", this is far from easy. We speak of attention as being involved in a huge variety of tasks: performing demanding tasks in distracting environments, orienting to sudden sounds, tracking the movement of multiple objects, noticing that an image has changed, or trying to keep one's mind clear of thoughts during meditation are all instances where it would be natural to speak of attention being used.

Given this diversity, it is not surprising that some theorists are skeptical that attention is a single process; as Allport (1993) puts it, "even a brief survey of the heterogeneity and functional separability of different components of spatial and nonspatial attentional control prompts the conclusion that, qua causal mechanism, there can be no such thing as attention." Two decades after Allport raised his doubts, there is still no clear consensus on whether there is a unifying neural or computational mechanism for attention (see Wu, 2014: 45-75 for a review).

This does not pose an immediate threat for defenders of the idea that attention is constitutively connected with the functions of Central Cognition. After all, even if there is no single neural mechanism for attention, it is possible that attention can be encompassed via a broad functional definition (such as the notion of "selection for task" in Wu, 2014: 38, or "cognitive unison" as defended by Mole, 2011) that places it squarely within the domain of activities performed by Central Cognition.

More worrying for defenders of the idea that attention is a function of Central Cognition is evidence that *visuospatial attention* and *central attention* are distinct psychological systems that do not interfere with one another (Fougnie, 2008). For example, one set of experiments (Pashler, 1989; 1991) studied attention by varying the temporal overlap between two attentionally-demanding tasks, one involving reaction times and the other involving visual identification tasks. If attention had a single capacity-limited focus (as argued by Cowan, 2001) or relied on a single controller (as in Baddeley, 1986), one would expect interference between the two tasks. However, Pashler found little effect on accuracy arising from task overlap, and concluded that "[e]mpirically… "attentional" processes seem to involve various dissociable mechanisms" (Pashler, 1994).

A second reason for doubting that attention is fundamentally connected to Central Cognition comes from the convoluted data concerning capacity limits in visual attention. Initial data on multiple object tracking and enumeration (e.g., Pylyshyn 1989; Trick & Pylyshyn 1993) suggested that subjects struggle to track or enumerate more than four objects at a time. The fact that this number also constitutes a limit for many tasks involving Central Cognition led theorists (notably Cowan, 2001) to suggest that performance in these experiments is limited by the resources of Central Cognition. However, subsequent work by Davis et al. (2001) suggested that the capacity limits of tracking and enumeration depend on overall stimulus *complexity* rather than the *number* of items (Fig. A0). The experiment of Davis et al. employed a task in which subjects were initially presented with either three larger objects or six smaller objects. After initial presentation, two notches (either jagged or square) briefly appeared on objects in the display. After the removal of the stimulus, subjects were required to say whether the notches were the same or different.

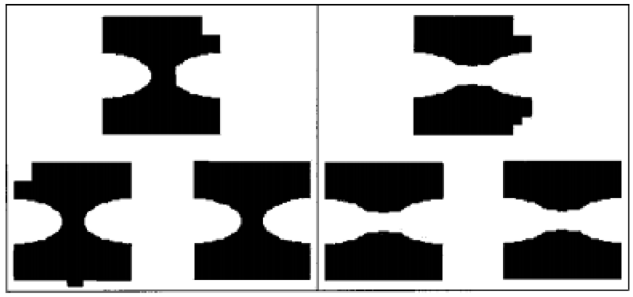


**Fig. A0.** 3- and 6-item displays like those used by Davis et al. (2001). Adapted from Davis (2001).

If visual attention was strictly limited to a number discreet objects, as Cowan has suggested, one would expect performance to be worse on the six object trials, since only four objects could be attended at a time. However, summarizing the data Davis notes that "[p]erformance was indistinguishable for the six- versus three object displays, suggesting that no four-object limit had operated, and that six objects can be attended as efficiently as three objects, once other display variables are equated" (Davis, 2001). Davis himself suggests a model in which the mechanisms of visual attention are subject to capacity limits arising from the overall *visual area* to be monitored and the total number of *features* to be tracked (in this case, the two notches) rather than the number of distinct *objects* presented.

More recent work on multiple object tracking (Alvarez & Cavanagh, 2005) within and across visual hemispheres provides further evidence against the idea that the limit of four items previously observed reflects a single centralized four-item capacity. Instead, Alvarez and Cavanagh found multiple-object tracking performance doubled from two to four when stimuli were spread across visual hemispheres rather than being confined to just one, leading them to claim that "the magic number is not four, but two plus two", this limitation arising not via Central Cognition but via a "matching bandwidth" in earlier perceptual processing.

Taken together, this evidence supports a dissociation between Central Cognition and at least some of the mechanisms of visual attention. To this extent, it suggest that models (like

Baddeley's and Cowan's) that aim to explain all attentional limitations purely via capacity limits in Central Cognition are overambitious. However, it would be a mistake to suppose that this cuts to the heart of the broader Sensory-Cognitive Model or the concept of capacity-limited Central Cognition. What I take to be crucial to the notion of Central Cognition is that it constitutes an important bottleneck in cognitive processing, imposing a limit on how many independently-chunked (or otherwise encoded) items can ultimately be reported or otherwise deployed for voluntary behavior at a given time. This is compatible with the existence of lower-level sensory bottlenecks that may have stricter and more localized capacity limits (as in Alvarez & Cavanagh 2005) or which are sensitive to quite different constraints on capacity such as feature complexity (as observed in Davis et al. 2001).

There is a great deal about attention that is still poorly understood, and as the small sample of experiments just reviewed illustrates, there are numerous distinctions to be drawn between stages of attention selection. For this reason, I have avoided discussing attention in the remainder of this dissertation save where absolutely necessary.

## Appendix 2: Variable Resource Models of Working Memory

In this dissertation I have made reference (notably in Chapter 1 and Chapter 2) to the observed four-item limit of working memory (Cowan, 2001). One natural way to interpret this limit is in terms of the idea that working memory has a limited number of 'slots', that is, can keep track of only four discreet items at a time once other factors are controlled for.

However, this simple slot-based model of Working Memory may be inadequate to explain a body of new data. Instead of taking the form of discrete slots, each of which may or may not be occupied by a given item, the four-item limit that is observed may be an artefact of a more flexible storage process in which representational resources can be allocated in different ways among different items, although there are several proposals for how this might work (see Fig A1).

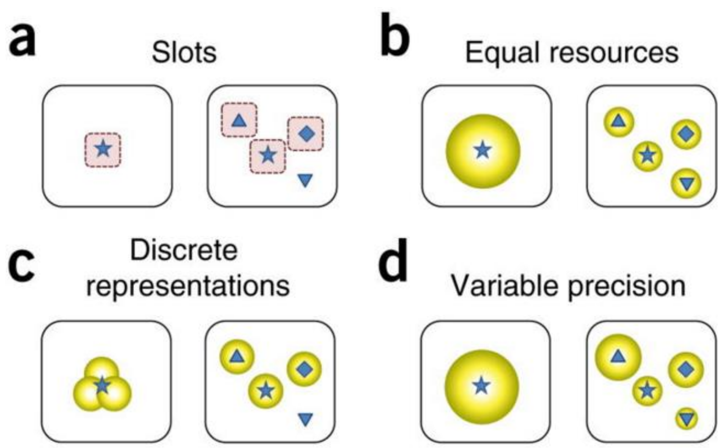


Fig. A1. Four models of resource allocation in Central Cognition (from Ma et al., 2014). The first model (a) is the traditional slot based model according to which subjects possess a limited number of memory slots, which can be occupied or unoccupied. The equal resource model (b) assumes that Central Cognition is a finite but continuous processing resource that can either be concentrated on a single item or spread (evenly) among multiple representations. The (c) discrete representations model combines elements of the slot- and resource-based models, and claims that the 'slots' in Central Cognition are discreet or 'quantized' information-maintenance channels that can nonetheless be combined or superimposed on a single item to improve accuracy of retention. Finally, the variable precision model (d) is similar to the equal resources model in assuming a continuous memory resource, but allows that items can be assigned varying rather than equal quantities of this resource in accordance with either top-down factors such as voluntary attention or bottom-up factors such as object salience.

What these new approaches tend to have in common is a rejection of the idea that storage in visual Central Cognition is simply a yes-or-no matter in which items are stored or not stored. Instead they claim items can be stored with varying degrees of precision. A simple proposal (the Discrete Representations position in Fig 2a) maintains the idea that there are slots in Central Cognition, but allows that multiple slots can be assigned to single items to improve overall precision of recall (see Zhang & Luck, 2008). Another more radical proposal claims that the notion of slots in Central Cognition should be abandoned altogether in favor of an entirely variable model in which memory limits are a function of the amount of 'noise' present in a given representation (Ma et al., 2014: 347). According to such models, noise levels can be affected by several factors, including cuing, but most importantly by the number of items that must be simultaneously retained at a given time.

In their emphasis on the fallible extraction of signals from noisy representations, these sorts of variable accounts are a natural fit for signal detection theories and the increasingly influential Bayesian models of visual perception. However, there is also more direct evidence in support of these approaches from experimental data that uses continuous measures of subjects' memory rather than discreet ones. For example, rather than asking a subject to report on the precise alphanumeric identity of a given letter or number, a subject might be asked to match the color or orientation of a presented stimulus using a variable slider or color wheel. This allows experimentalists to assess in more detail how well subjects remembered presented stimuli. These paradigms have produced some results that are surprising for defenders of slot-based models. For example, it seems that at least some errors in demanding working memory tasks increase smoothly rather than discreetly in proportion to the number of items presented (see Fig A2). That is, rather than subjects exhibiting flawless performance up to the magic number four, and massively degraded performance thereafter, their performance seems to be best for single items and decrease with each additional item that must be recalled. Additionally, there seems to be considerable variation in the reliability of recall for individual items within subjects in some contexts: properties of items that are cued in advance are remembered more accurately than items that are not cued in cued trials, or items in general in uncued trials (see Fig A3).

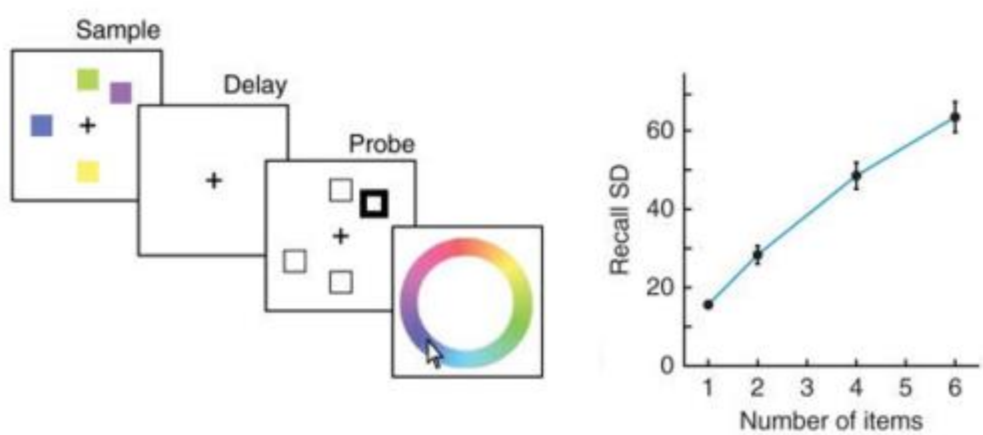


Fig. A2. A color delayed-estimation task (from Ma et al., 2014, summarizing Wilken & Ma, 2004). Subjects were presented with a variable number of colored squares, then given a subsequent cue after a short delay. Their task was to indicate on a color wheel the precise color of the cued square. Their accuracy (as measured by increasing recall variability) decreased smoothly as a function of the number of squares in the initial sample. This is in contrast to the slot-based model, which would predict a flat performance up to four items and a rapidly degrading performance thereafter.

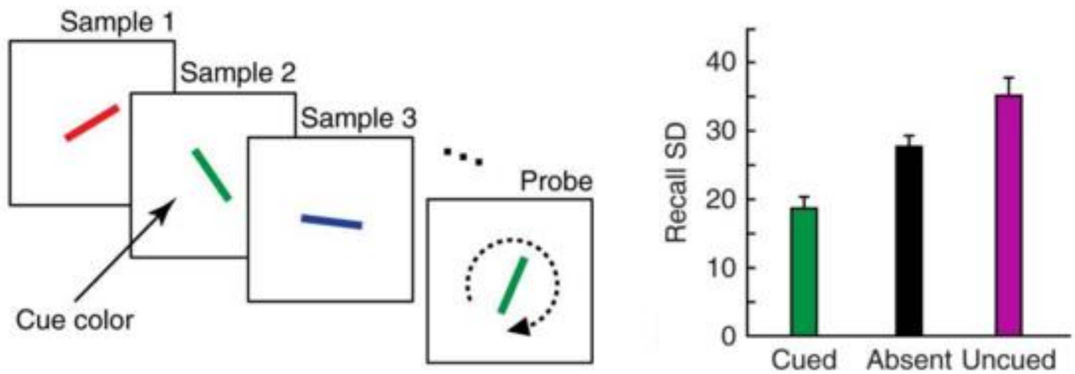



**Fig. A3.** An orientation delayed-estimation task (from Ma et al., 2014, summarizing Gorgoraptis et al. 2011). Subjects were sequentially presented with colored lines at various angles. Their task was to indicate the orientation of probed lines using a response dial. In some trials, subjects were told in advance that a given color was likely to be cued. The recall variability of subjects' responses was significantly higher for uncued items relative to cued items. Even more interestingly, however, the recall variability of subjects' responses in trials without any cues was intermediate between recall variability of cued and uncued items in trials using cues. This suggests that subjects can variably allocate representational resources in accordance with task demands, rather than simply retaining or not retaining representations as might be predicted by a simple slot-based model.

Do these accounts constitute a potentially fatal threat to the Sensory-Cognitive Model described in Chapter 1 or the amended version presented in Chapter 2? I do not believe so. The Sensory-Cognitive Model is defined not by a commitment to the idea that items in memory are stored in single slots, but rather to the idea that short-term memory can be understood in terms of the interaction between two kinds of memory faculty, namely high-capacity, brief duration sensory stores and a more limited-capacity Central Cognition, plus, as argued in Chapter 2, a further intermediate store in the form of CSTM. Even if, as advocates of variable models maintain, visual Central Cognition should not be understood in terms of a finite number of slots, the data continues to support the idea that for many tasks there is a fairly stable *functional* limit as to the maximum number of items subjects can retrieve with any accuracy (indeed, Bays et al., 2011 describe such a limit in the course of arguing for a variable resource account of Central Cognition). Additionally, it is worth noting that Cowan himself seems to regard current debates about variable mechanisms of storage not as a threat to his broader work but simply as evidence of the growing sophistication of the kind of model he advocates (Logie & Cowan, 2015).

That being said, some theorists, notably Gross & Flombaum (forthcoming), have argued that the variable resource model might tell against the idea of perception as involving a series of memory stores of declining capacity, a claim that is more directly in tension with the Sensory-Cognitive Model. For example, they suggest that partial report superiority (as observed in the Sperling task) can be entirely explained in terms of the precision-enhancing effects of cues on rich but noisy representations in Central Cognition, in an account that denies any direct role for sensory memory stores.

Gross and Flombaum's project is ambitious, and it remains unclear whether their account can satisfactorily describe certain distinctive features of partial report superiority (for example, its vulnerability to interference from pattern masks, a feature that is not shared by encoding in Central Cognition). Moreover, even if Gross and Flombaum are correct in suggesting that these particular experiments can be explained without iconic memory specifically, there is a host of independent evidence (reviewed in Chapter 1) to support a broader functional dissociation between Central Cognition and the various forms of sensory memory. Hence while variable resource models of Working Memory may provide the best way forward in understand encoding in Working Memory, they need not displace the existing distinction between sensory and cognitive forms of short-term memory.

## Appendix 3: Unconscious Working Memory

Throughout this dissertation, I have assumed that information in working memory is conscious and available for report. In this respect, I have been following a common assumption made by leading theorists in the field (Baars and Franklin, 2003, Baddeley, 2003, Cowan, 2001). However, very recent work has suggested that there may be a form of working memory that allow for information to be stored and even subsequently used by a subject without becoming conscious.

The majority of trials proceed as follows (see Fig. A4). First, subjects are shown a low visibility stimulus such as a Gabor patch that is immediately masked. After a variable delay (ranging from a few hundred milliseconds to up to 15 seconds) subjects are then presented with a visible stimulu and asked to indicate whether it is rotated clockwise or anticlockwise relative to the originally presented item. A number of studies (Bergström & Eriksson, 2014, 2015; Dutta et al., 2014; Soto et al., 2011) have found that subjects still perform significantly above chance in the memory task, despite reporting not having seen the initial stimulus. This has been taken to constitute evidence for unconscious *working* memory specifically because the significant delay times and use of masking would seem to rule out other forms of memory such as sensory memory (and perhaps CSTM).

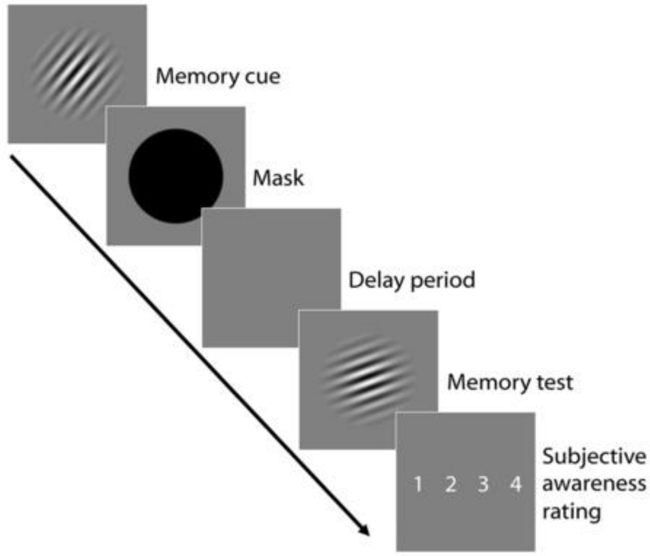



**Fig. A4. A sample paradigm from experiments on working memory** (from Stein, Kaiser, & Hesselmann, 2016; cf Soto et al. 2011).

Two main lines of criticism have been levied against interpreting this data as showing evidence of unconscious working memory, however. First, it seems an open possibility that subjects, despite reporting not having seen the initial stimulus did, in fact undergo brief or degraded conscious experience of the original stimulus (Phillips. 2015; see Chapter 4.3). While the possibility of unreportable conscious experience is sometimes used to attack working memory theories of consciousness, it also seems possible that, in some cases information can be consciously encoded in working memory in a highly degraded form such that they are non-confident of having seen anything, even if it can influence their subsequent judgments (Schmidt, 2015).

A second possibility is that subjects undergo an unconscious *perception* when presented with the initial cue, and immediately make a conscious *guess* about the features of the initial memory cue that is influenced by their unconscious perception and subsequently stored in working memory. This conscious representation might then explain their above-chance performance on the memory task. In order to rule out this possibility, it would need to be demonstrated that subjects are at chance when asked to make immediate guesses about the orientation of the memory cue. However, there is arguably no empirical work that clearly demonstrates this possibility (Stein, Kaiser, & Hesselmann, 2016).

One recent set of experiments on unconscious working memory (King, Pescetelli, & Dehaene, 2016) may provide a better insight into the neural underpinnings of some unconscious working memory tasks. King et al. used a paradigm broadly similar to that illustrated in Fig. 4A above, and subjects were asked questions about the presence, orientation, and visibility of the the stimulus. King et al. found, as in previous work, that subjects were able to make above-chance guesses about the orientation of the stimulus in question, even when they reported that it was not visible and had not seen it. Importantly, however, this experiment used MEG and sophisticated forms of data analysis to decode the neural underpinnings of subjects' memory abilities. Specifically, it was found that early (<250ms) neural activity encoded all of the features of the stimulus, while later activity (>300ms) encoded only the *task relevant* features of the stimulus, namely its presence, orientation, and visibility, suggesting that maintenance of information was affected by the specific demands of the task at hand. King et al. suggests this shows a dissociation between "automatic encoding and the selective maintenance of visual features." This latter information, the authors suggest, can be subsequently partially broadcast "across the cortical hierarchy even when the stimulus remains subjectively invisible", and subsequently recruited for task performance.

This trial may suggest that at least some forms of unconscious working memory are not best explained in terms of unconscious perception combined with conscious guesswork. If that were the case, then one might anticipate brief activations in sensory areas coupled with normal activity in the global neuronal workspace. The kind of weak 'partial broadcast' observed instead suggests that unconscious working memory may reflect a distinctive kind of psychological process that is not mediated by normal working memory.

Of course, even if this information was not encoded in working memory or in sensory areas, it remains possible that it was conscious, despite subjects' reports to the contrary. In light of the broader discussions in this dissertation (especially Chapter 4), we might even speculate as to whether CSTM was involved. I think it unlikely that CSTM can help us explain other unconscious working memory paradigms, given the extended intervals (up to 15 seconds) between the cue and test stimuli. However, one critical feature of this experiment was that the delay time between the memory cue and the memory test was extremely short: just 800ms. This is very similar to the delays used by Potter (2014 et al.). Of course, this experiment involved comparisons across fairly low level features such as visibility and orientation, marking a clear difference from Potter's work. However, note that this form of unconscious working memory does not seem to be based in lower-level sensory areas, and moreover involves apparently active selection just of task-relevant features for retention. This may in turn suggest some degree of interface between a subject's knowledge of the task demands and the selection of which information is to be retained for longer than 300ms. It therefore seems a possibility worth exploring that the neural processes involved in this form of unconscious working memory may be relevantly similar to those involved in CSTM.

## Appendix 4: Challenging the Tripartite Scheme

I suggest in Chapters 1 and 4 that we can distinguish many leading theories of consciousness in terms of how they spell out the relationship between consciousness and cognitive access, thereby distinguishing between access, access-independent, and accessibility accounts. While this tripartite scheme is not intended to be exhaustive, it is worth briefly exploring why this may fail to properly accommodate certain accounts.

First, note that many leading theories do not make specific commitments about cognitive access. The last two decades have witnessed an explosion in the variety of theories of consciousness, with diverse starting points such as mathematical models of information integration (Tononi, 2008), the success of predictive models in neural networks (Clark, 2016; Hohwy, 2013), and the dynamic relationship between action and perception (Noe, 2005). Many of these theories engage only indirectly with questions about the role of short-term memory processes and cognitive access. Given the nuances and differences among such theories, it would therefore be overambitious to attempt to shoehorn them into the tripartite model provided.

However, a few key questions can be addressed to these outlying theories that enable us to make some sort of sense of them in light of the tripartite framework. For example, we can ask of any given theory whether it claims that a constitutive feature of conscious states is that subjects be able to deploy them in voluntary action such as report. Any theory that answers in the negative can be reasonably considered to be broadly speaking an access-independent account.

To illustrate, consider Tononi's Integrated Information Theory, or IIT (Tononi & Koch, 2015). Without going into fine details, IIT takes consciousness to be a property of integrated systems, and to exist among a system's states at a given time to the extent that the system exhibits a high degree of informational interconnectivity (measured by a mathematically defined concept termed Phi). The references to interconnectivity might misleadingly suggest that IIT would be committed to some minimal cognitive accessibility constraint. However, Tononi makes clear that consciousness might arise to a limited degree in systems that lack any central cognitive processing, including very simple systems like photodiodes (Tononi, 2008). While this does not automatically commit him to view that there are, in fact, inaccessible conscious states in human beings (though other defenders of IIT are more explicit in this regard; see Koch & Tsuchiya, 2007) it nonetheless suggests that IIT might be very loosely grouped with access-independent theories.

A second kind of worry cuts deeper for the tripartite distinction introduced thus far. Specifically, one might question whether cognitive access is really a single process, or instead misleadingly lumps together a number of distinct and separable forms of executive cognition.

This important objection certainly highlights a neglected aspect of Block's distinction between phenomenal consciousness and cognitive access, namely his normally tacit assumption that there is a fairly unified set of mechanisms underlying cognitive access. While we could perhaps still make some sense of the notion of cognitive access even if the mind has a massively distributed and fragmented architecture, this is clearly not what most access and Accessibility theorists have in mind. Instead, they take cognitive access to describe a psychological natural kind involving the global broadcast of information.

If this psychological kind turned out not to exist, then it would not follow that Block's distinction was no longer relevant. However, it might force us to make some further distinctions, specifying, for example, which cognitive systems were constitutively connected to consciousness and which were not. As matters stand, however, I would suggest that the psychological and neural evidence for Global Workspace Theory points to a broadly unified set of cognitive mechanisms that are involved in high-level activities such as report (although see Appendix 3, above).

A related objection might grant the existence of the global workspace while insisting that there was some further cognitive process that was required for consciousness. I have in mind here, for example, Rosenthal's higher-order thought view, which claims that a state becomes conscious when, via a suitable process, it becomes the target of some higher-order thought (Rosenthal, 2005). Crucially, Rosenthal claims that mechanisms of metacognition are separable from mechanisms of global broadcast, allowing for states to be globally broadcast without being the targets of actual higher-order thoughts (thus, on his view, remaining unconscious), and equally, for states to be the targets of higher-order thoughts without being globally broadcast (thereby being conscious without being globally available; Rosenthal, 2009).

I do not know of any specific neuroscientific evidence for these claims, but if this view is correct, then it is possible for consciousness to occur without cognitive access in the sense of global broadcast but not without cognitive access *tout court* (specifically, it requires meta-cognition). This offers a further complication for the tripartite scheme described earlier. The best that can be said at this point is that, though the details are complex, Higher-Order Thought theory bears at least important resemblances to leading access accounts of consciousness. Rosenthal (2005) claims, for example, that a subject's possessing a higher-order thought suffices to make the relevant first-order state reportable. If this is right, then the higher-order theorist and the access theorist will agree about most cases of conscious experience, and the vital role of report in pinning down the presence or absence of conscious states. However, this serves as another reminder that the tripartite distinction provided has its limitations.

# Appendix 5: Apperceptive and Associative Agnosia

In Chapter 3, I mention associative agnosia as an example of a condition in which subjects seemingly lose their capacity for categorical perception despite possessing intact Central Cognition. In this appendix, I will briefly explore this condition further, and spell out in a little more detail how it might be explicable in terms of the CSTM account of categorical perception.

Broadly speaking, visual agnosia involves a subject's inability to recognize an object when it is presented visually, even though the subject often has largely preserved cognitive skills and memory and is often able to recognize stimuli in other sensory modalities (for example, by touch). Note that, while visual agnosias are often diagnosed based on subjects' inability to *name* objects, the condition as a whole goes beyond mere anomia, since agnosic patients typically cannot even identify, for example, suitable uses for a given object or say what other objects it is associated with.

A major distinction can be drawn between two kinds of visual agnosia, namely *apperceptive* and *associative* agnosia (a distinction suggested by Lissauer, 1890). While both apperceptive and associative agnosics suffer from impaired object recognition, the former seem to have drastically impaired visual information about stimuli, while the latter have lost access just to semantic information (Gottfried, 2011: Ch.11). This can be assessed via the simple procedure of asking a subject to copy a picture of an object. Although they can name individual features of presented images, such as colors and lines, apperceptive agnosics are unable to copy whole pictures (Rubens & Benson, 1971). By contrast, associative agnosics will generally be able to copy the image, although, tellingly, they tend to copy it 'slavishly', preserving distortions or inaccuracies in the original that normal subjects would correct (Farah, 2004; see Fig. A5). Rubens and Benson describe one such patient as follows.

> Ability to recognize pictures of objects was greatly impaired, and after repeated testing he could name only or two out of ten line drawings. He was able to name geometrical forms (circle, square, triangle, cube). Remarkably, he could make excellent copies of line drawings and still fail to name the subject… he easily matched drawings of objects that he could not identify, and had no difficulty discriminating between complex nonrepresentational patterns differing from each other only subtly. He occasionally failed in discriminating because he included imperfections in the paper or in the printer's ink. (Rubens & Benson, 1971)

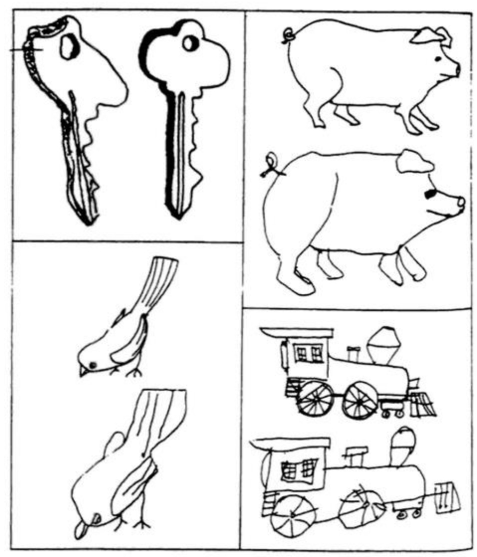


**Fig. A5.** Examples of images that an associative agnosic subject could not name, together with their 'slavish' copies. From Rubens & Benson, 1971.

If the proposals about categorical perception given in Chapter 3 are along the right lines, the distinction between these apperceptive agnosia might naturally be spelled out in terms of a distinction between high-level perceptual processing and the kind of truly categorical perception underpinned by CSTM. Note first that whereas apperceptive agnosia seems to be caused by damage to sensory areas, associative agnosia is linked to damage to medial temporal areas (Greene, 2005). Apperceptive agnosics, then, may have impaired *perception*, meaning that while they can detect color and individual features, they are unable to resolve what they are seeing into coherent wholes.

By contrast, the fact that subjects with associative agnosia have no difficulties matching images just on the basis of their visual form and can copy them adequately suggests that they do not have this kind of perceptual deficit. Similarly, the fact that the patient described above could recognize basic geometric shapes suggests that what is impaired is not visual recognition *tout court* but rather the ability to *classify* perceptions in accordance with higher-level object categories. The fact that associative agnosics also exhibit intact central cognitive processing (Gottfried, 2011: Ch.11) suggests that the deficit is also not an impairment of Central Cognition. These considerations suggest that the deficit is neither strictly perceptual nor cognitive, but constitutes an impairment to an intermediate faculty, namely CSTM.

I take this hypothesis to be broadly congruent with the more specific neurological hypotheses advanced to explain the condition. For example, Teuber (1968) and Warrington (1975) offer accounts of associative agnosia that posit an impairment in the mechanisms linking perception to stored knowledge of objects (indeed, Teuber talks of associative agnosics as experiencing a “percept stripped of its meaning”). This is broadly the function I take CSTM to fulfill, namely connecting incoming perceptual information to semantic categories and making the subsequent representations available to Central Cognition.

It is worth considering at least one opposing hypothesis, however, namely that associative agnosics are deficient not in terms of a stage of processing but in terms of *stored semantic memory*; in other words, that they have all the same faculties they had before, but have lost one store of information called upon by such faculties. Thus it might be suggested that associative agnosics have suffered damage to a stored 'database' of images paired to semantic descriptions, such that central cognitive processes could no longer classify images in respect of their specific semantic identity.

Such an explanation could account for associative agnosics' impairments without positing an independent faculty bridging perception and cognition. However, it also stands at odds with more specific details regarding associative agnosic patients. For example, some agnosics are able to report on the visual details of objects from memory (Behrmann, Winocur, & Moscovitch, 1992), suggesting that their knowledge of the visual attributes of objects is intact. Even more strikingly, Jankowiak et al. (1992) report a subject who "remained able to draw complex objects from memory but could not subsequently recognize his sketches." Again, this suggests that the deficit is not one of knowledge, but rather of function, and specifically the ability to fluently categorize incoming perceptual information in light of stored information.

In the context of connecting models of perception and cognition with cognitive neuroscientific data, the evidence from associative agnosics further strengthens the case for positing an intermediate mechanism linking the two. It is, of course, a further and more controversial claim that CSTM is the precise mechanism involved. However, as current data stands, there is to my knowledge no candidate form of short-term processing mechanism that provides a better match for the psychological and neuroscientific data.

## Appendix 6: Dehaene's Signatures of Consciousness

In Chapter 1 and Chapter 4, I briefly summarize the Global Neuronal Workspace theory of Stanislas Dehaene. In this brief appendix, I will spell out a few more details of this view. I do so in part because Dehaene's account of consciousness is impressively grounded in empirical research and worth considering in a little more detail. However, I also wish to make strengthen one part of the case offered in Chapter 4 that there are genuinely two systems underlying conscious experience, one being CSTM, and the other being the kind of Global Neuronal Workspace spelled out by Dehaene.

Dehaene's model falls within the family of theories known as Global Workspace accounts. Thus putting his view succinctly, Dehaene states that "[w]hen we say that we are aware of a certain piece of information, what we mean is just this: the information has entered into a specific storage area that makes it available to the rest of the brain" (2014). In this respect, Dehaene's view builds fairly straightforwardly on the framework of Baars, but Dehaene has gone further in providing empirical evidence of just such an information sharing system in the form of a network of dense pyramidal neurons concentrated in the frontal areas of the brain but providing links to a wide range of other areas (Dehaene, Changeux, Naccache, 2011).

Whenever we have a conscious experience, Dehaene claims, this network of neurons is activated in a multi-step process which he breaks down into four "signatures of consciousness". The first step consists in a winner-takes-all process in which 'coalitions' of sensory neurons encoding similar features (cf. Koch and Crick 2003) cross a critical threshold of firing and trigger a 'global ignition', a widespread set of firings in a prefrontal and parietal areas that are relatively silent prior to subjects' becoming aware of a stimulus (Naccache & Dehaene 2001). This wave of firings produces a second signature of consciousness known as the ‘P3 wave’, a slow and substantial spike in positive voltage across the top of the head that Dehaene claims is absent in cases of unconscious perception. A third signature of conscious thought suggested by Dehaene (but defended by others, including Prinz, 2012) is a massive increase in high frequency (>30Hz) oscillations in neurons, starting around the same time as the P3 wave (300ms after stimulus detection). Finally, in a fourth signature of consciousness, we see these oscillations synchronize between frontal and sensory areas of the brain, leading to strikingly uniform gamma-band oscillations synchronized across the brain.

Via an expansive research program, Dehaene has accumulated data that suggests these four neural signatures of conscious experience are presented in all and only cases in which subjects can report on a given stimulus; they are not present in those instances where a subject perceives a stimulus subliminally, even though other forms of firing might be present (for a review, see Dehaene 2014). Buttressing this impressive neuroimaging data is further evidence from cognitive neuroscience. For example, lesions to areas of the brain containing especially dense clusters of the pyramidal neurons involved in long-distance communication in the relays seem to cause neglect, or loss of cognitive access to stimuli (Dehaene, 2014: 170). Moreover, the extent to which the global neuronal network remains present in patients in persistent vegetative states serves as a reliable predictor of patients’ likelihood of recovery (see Fig. A6).

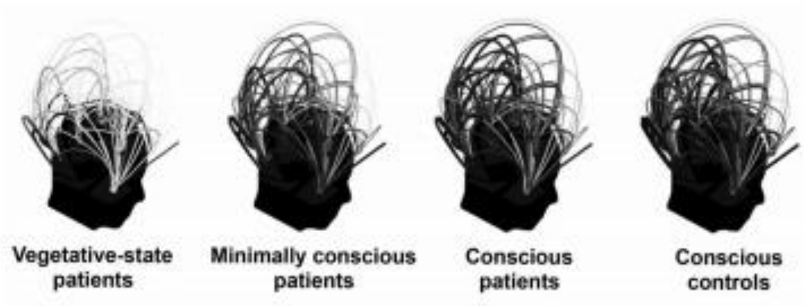



**Fig. A6. A** comparison of the degree of information exchanged across long-cortical distances as measured by EEG in a group of almost 200 persistent vegetative state patients and controls. The thickness of lines indicates the amount of information shared across different EEG sites. Those vegetative state patients showing high degrees of information sharing were more likely to regain full consciousness in the following weeks and months. See Dehaene, 2014.

We have good reason, then, for thinking that Dehaene has isolated an important component of conscious processing, and identified its neural signatures. As noted in Chapter 4, I think that this kind of mechanism may be responsible for the kind of attentionally-demanding conscious processing that underpins things like complex visual search tasks or reflective thought. However, as I argued there, I suggest that this is just one mode of consciousness, the other being provided by CSTM.

# REFERENCES


- Allen, C. & Hauser, M. (1996). Concept Attribution in Nonhuman Animals: Theoretical and Methodological Problems in Ascribing Complex Mental Processes. In M. Bekoff & D. Jamieson (eds.), *Readings in Animal Cognition*. Cambridge MA, MIT Press. 47–62.
- Allen, C., Grau, J., Meagher, M.W. (2009). The Lower Bounds of Cognition: What Do Spinal Cords Reveal? In Bickle, John (ed.) *The Oxford Handbook of Philosophy and Neuroscience*. Oxford University Press.
- Allport, A. (1993). Attention and Control. Have we been asking the wrong questions? A critical review of twenty-five years. In D.E. Meyer & S.Kornblum (eds.), *Attention and Performance* XIV. The MIT Press 183-218.
- Alvarez, G. A., & Cavanagh, P. (2005). Independent Resources for Attentional Tracking in the Left and Right Visual Hemifields. *Psychological Science, 16*(8), 637-643.
- Andrade, J. (2001). *Working memory in perspective*. Hove, East Sussex: Psychology Press.
- Atkinson, R., & Shiffrin, R. (1968). Human Memory: A Proposed System and its Control Processes. *Psychology of Learning and Motivation,* 89-195.
- Avargues-Weber, A., Dyer, A. G., & Giurfa, M. (2010). Conceptualization of above and below relationships by an insect. *Proceedings of the Royal Society B: Biological Sciences, 278*(1707), 898-905.
- Averbach, E., & Coriell, A. S. (1961). Short-Term Memory in Vision. *Bell System Technical Journal*, *40*(1), 309-328.
- Baars, B. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences*, 6: 47–52.
- Baars, B. J., & Franklin, S. (2003). How conscious experience and working memory interact. *Trends in Cognitive Sciences, 7*(4), 166-172.
- Baars, Bernard J. (1988). *A Cognitive Theory of Consciousness*. Cambridge University Press.
- Baddeley, A. (1992). Consciousness and working memory. *Consciousness and Cognition, 1*(1), 3-6.
- Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.
- Baddeley, A.D. & Hitch, G.J. (1974) Working memory. In G.Bower (Ed.), *The Psychology of learning and motivation* (pp. 47–89). New York: Academic Press
- Baddeley, A.D. (1986) *Working memory*. Oxford: Oxford University Press.
- Baddeley, A.D. (2000). The episodic buffer: A new component of working memory?*Trends in Cognitive Sciences, 4*(11), 417-423.
- Baddeley, A.D. (2003). Working memory: Looking back and looking forward.*Nature Reviews Neuroscience*, *4*(10), 829-839.
- Baddeley, A.D., Lewis, V.J. & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, 36A, 233–252
- Baddeley, A.D., Thomson, N. & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14, 575– 589.
- Barron, A. B., & Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proceedings of the National Academy of Sciences Proc Natl Acad Sci USA, 113*(18), 4900-4908
- Bayne, T. (2007). Conscious states and conscious creatures: Explanation in the scientific study of consciousness. *Philosophical Perspectives* 21 (1):1–22.
- Bayne, T. (2009). Perception and the reach of phenomenal content. *Philosophical Quarterly* 59 (236):385-404.

- Bayne, T. (2010). *The unity of consciousness*. Oxford: New York.
- Bayne, T., & Montague, M. (2011). Introduction to Cognitive Phenomenology. In Bayne, T., & Montague, M. (eds), *Cognitive Phenomenology*. Oxford University Press UK.
- Behrmann, M., Winocur, G., & Moscovitch, M. (1992). Dissociation between mental imagery and object recognition in a brain-damaged patient. *Nature*, 359(6396), 636-637.
- Belke, E., Humphreys, G., Watson, D., Meyer, A. and Telling, A., (2008). Top-down effects ofsemantic knowledge in visual search are modulated by cognitive but not perceptual load. *Perception & Psychophysics*, 70 8, 1444 – 1458.
- Bergström, F., & Eriksson, J. (2014). Maintenance of non-consciously presented information engages the prefrontal cortex. *Frontiers in Human Neuroscience* 8:938.
- Bergström, F., & Eriksson, J. (2015). The conjunction of non-consciously perceived object identity and spatial position can be retained during a visual short-term memory task. *Frontiers in Psychology* 6:1470
- Bermúdez, J. L. (2003). *Thinking without words*. Oxford: Oxford University Press.
- Block, N. (1995). On a Confusion about a Function of Consciousness, *Behavioral and Brain Sciences*, 18, pp. 227–287.
- Block, N. (1996). Mental Paint and Mental Latex. *Philosophical Issues, 7*, 19.
- Block, N. (2007). Consciousness, Accessibility, and the Mesh Between Psychology and Neuroscience, *Behavioral and Brain Sciences* 30, pp. 481–499.
- Block, N. (2011a). Perceptual Consciousness Overflows Cognitive Access, *Trends in Cognitive Science* 12, pp. 567–575.
- Block, N. (2011b). The Anna Karenina Theory of the Unconscious.*Neuropsychoanalysis, 13*(1), 34-37.
- Block, N. (2014a). Seeing-As in the Light of Vision Science. *Philosophy and Phenomenological Research, 89*(3), 560-572.
- Block, N. (2014b). Rich Conscious Attention Outside Focal Attention, *Trends in Cognitive Sciences* 18, pp. 445–447.
- Block, N. (2014c). Consciousness, Big Science and Conceptual Clarity. In G. Marcus & J. Freeman (eds.), *The Future of the Brain: Essays by the World's Leading Neuroscientists*. Princeton University Press 161-176.
- Block, N. (2015). Solely Generic Phenomenology. *Open MIND 2015*.
- Block, N. (forthcoming) Tweaking the Concepts of Perception and Cognition. *Behavioral and Brain Sciences.*
- Bloom, A.(1968) *The Republic of Plato*, Basic Books.
- Bratch, A., Kann, S., Cain, J., Wu, J., Rivera-Reyes, N., Dalecki, S, Crystal, J. (2016). Working Memory Systems in the Rat. *Current Biology, 26*(3), 351-355.
- Brewer, B. (2005). Perceptual experience has conceptual content. In Ernest Sosa & Matthias Steup (eds.), *Contemporary Debates in Epistemology*. Blackwell
- Broadbent, D.E. (1975). The magic number seven after fifteen years. IIn Kennedy, A., Wilkes, A. (eds). *Studies in Long-Term Memory.* Oxford, England: Wiley, pp. 3–18.
- Brooks, D. I., Freeman, J. H., & Wasserman, E. A. (2013). Categorization of photographic images by rats using shape-based image dimensions. *Journal of Experimental Psychology: Animal Behavior Processes, Vol 39(1), Jan 2013, 85-92*
- Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, 64, 123–152.
- Buijs, S., Keeling, L. J., & Tuyttens, F. A. (2011). Using motivation to feed as a way to assess the

importance of space for broiler chickens. *Animal Behaviour*, 81(1), 145-151.

- Burge, T. (2007). Psychology supports independence of phenomenal consciousness. *Behavioral and Brain Sciences* 30 (5-6):500-501.
- Burge, T. (2010a). Origins of Perception. *Disputatio* 4 (29):1 - 38.
- Burge, T. (2010b). *Origins of objectivity*. Oxford: Oxford University Press.
- Burge, T. (2014b). Perception: Where Mind Begins. *Philosophy* 89 (3):385-403.
- Burge, T. (2014a). Reply to Block: Adaptation and the Upper Border of Perception. *Philosophy and Phenomenological Research* 89 (3):573-583.
- Butler, A., Oruc, I., Fox, C., & Barton, J.(2008). Factors contributing to the adaptation after effects of facial expression. *Brain Research 1191*: 116–126
- Camp, E. (2007). Thinking with maps. *Philosophical Perspectives* 21 (1):145–182.
- Caramazza, A., & Shelton, J. R. (1998). Domain-Specific Knowledge Systems in the Brain: The Animate-Inanimate Distinction. *Journal of Cognitive Neuroscience,10*(1), 1-34.
- Carey, S. (2009). *The Origin of Concepts*. Oxford University Press.
- Carey, S. (2011). Précis of the origin of concepts. *Behavioral and Brain Sciences* 34 (3):113-124.
- Carruthers, P. (2005). *Consciousness: Essays From a Higher-Order Perspective*. Oxford University Press Uk.
- Carruthers, P. (2009). Invertebrate concepts confront the generality constraint (and win). In R.W. Lurz (ed.), *The Philosophy of Animal Minds*. Cambridge University Press 89-107.
- Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. OUP Oxford.
- Carruthers, P. (2014). On central cognition. *Philosophical Studies* 170 (1):143-162.
- Carruthers, P. (2015a). *The Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*. Oxford University Press Uk.
- Carruthers, P. (2015b). Block's Overflow Argument. *Pacific Philosophical Quarterly* 97 (2):n/a-n/a.
- Carruthers, P., & Veillet, B. (2011). The Case Against Cognitive Phenomenology. In Bayne, T., & Montague, M. (eds.), *Cognitive phenomenology*. Oxford University Press.
- Cha, M., Kohan, K. J., Zuo, X., Ling, J. X., & Gu, J. G. (2012). Assessment of chronic trigeminal neuropathic pain by the orofacial operant test in rats. *Behavioural Brain Research*, *234*(1), 82-90.
- Chalmers, David J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2 (3):200-19.
- Chater, N., & Heyes, C. (1994). Animal Concepts: Content and Discontent. *Mind & Language, 9*(3), 209-246.
- Chittka, L., & Jensen, K. (2011). Animal Cognition: Concepts from Apes to Bees.*Current Biology*, *21*(3)
- Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press Usa.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36 (3):181-204.
- Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*(12), 547-552.
- Corns, J. (2014). Unpleasantness, Motivational Oomph, and Painfulness. *Mind & Language* 29.2: 238-54.
- Corrigan, M. S. (2007). *Pattern recognition in biology*. New York: Nova Science.
- Cowan, N., (2001). The magical number 4 in short-term memory: A reconsideration of mental storage

capacity. *Behavioral and Brain Sciences 24*1, 87-114.

- Craik, F. I., & Watkins, M. J. (1973). The role of rehearsal in short-term memory.*Journal of Verbal Learning and Verbal Behavior*, *12*(6), 599-607.
- Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, *6*(2), 119-126.
- Dade, L., Zatorre, R., Evans, A., & Jones-Gottman, M. (2001). Working memory in another dimension: Functional imaging of human olfactory working memory. *Neuroimage*, 14, 650–660.
- Daltrozzo, J., Signoret, C., Tillmann, B., & Perrin, F. (2011). Subliminal Semantic Priming in Speech. *PLoS ONE*, *6*(5).
- Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, 32, 255-267.
- Davidson, D. (1975). Thought and Talk. In S. Guttenplan (ed.), *Mind and Language*. Oxford: Oxford University Press. 7–23.
- Davis, G. (2001). There is no four-object limit on attention. *Behavioral and Brain Sciences*, *24*(1), 119-120.
- Davis, G., Welch, V. L., Holmes, A., & Shepherd, A. (2001). Can attention select only a fixed number of objects at a time? *Perception*, *30*(10), 1227-1248
- Dawkins, M. S. (2012). *Why animals matter: Animal consciousness, animal welfare, and human well-being*. New York: Oxford University Press.
- De Busscher, J., Hoffman, G., and Kluyskens, J. (1956). Agnosia visuelle temporaire pour les personnes et opticospatiale pour les objets a la suite d'un ictus unique. *Acta Neurologica Belgica*, 56, 162-176.
- Dehaene, S. (2016). The Relationship Between Working Memory And Consciousness: A Reexamination Using MEG Decoding Techniques. Presentation at the 20th meeting of the Association for the Scientific Study of Consciousness
- Dehaene, S., & Naccache, L., (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, 79:1-37.
- Dehaene, S. (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking Press, 2014.
- Dehaene, S., Changeux, J., & Naccache, L. (2011). The Global Neuronal Workspace Model of Conscious Access: From Neuronal Architectures to Clinical Applications. *Research and Perspectives in Neurosciences Characterizing Consciousness: From Cognition to the Clinic?,* 55-84.
- Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., et al. (1998). Imaging unconscious semantic priming. *Nature*, 395, 597–600.
- den Daas, C., Häfner, M., & de Wit, J. (2013). Sizing opportunity: Biases in estimates of goal-relevant objects depend on goal congruence. *Social Psychological and Personality Science*, 4, 362–368
- Dennett, D.C. (1993).The message is:There is no medium. *Philosophy and Phenomenological Research* 53:919–31.
- Dennett, D.C. (1996). *Kinds of minds: Toward an understanding of consciousness*. New York, NY: Basic Books.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. MIT Press.
- Dutta, A., Shah, K., Silvanto, J., & Soto, D. (2014). Neural basis of non-conscious visual working memory. *Neuroimage* 91:336–43.
- Egan, A. (2008). Seeing and believing: perception, belief formation and the divided mind. *Philosophical Studies* 140 (1):47 - 63.
- Elwood, R. W., & Appel, M. (2009). Pain experience in hermit crabs? *Animal Behaviour, 77*(5), 1243-

1246.

- Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory.*Journal of Experimental Psychology: General*, *143*(2), 548-565.
- Enns, J. T., & Di Lollo, V. (1997). Object substitution: A new form of masking in unattended visual locations. *Psychological Science*, 8, 135-139
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245.
- Ericsson, K., Delaney, P. F., Weaver, G., & Mahadevan, R. (2004). Uncovering the structure of a memorist's superior basic memory capacity. *Cognitive Psychology*, 49(3), 191-237.
- Farah, M. J. (2004). *Visual agnosia*. Cambridge, MA: MIT Press.
- Firestone, C. (2013). On the origin and status of the El Greco fallacy. *Perception*, 42, 672–674.
- Firestone, C., & Scholl, B.J. (2015). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behavioral and Brain Sciences*:1-77.
- Fodor, J. (1983). *Modularity of Mind*. MIT Press.
- Fodor, J. (2007). The revenge of the given. In Brian P. McLaughlin & Jonathan D. Cohen (eds.), *Contemporary Debates in Philosophy of Mind*. Blackwell 105--116.
- Fodor, J. (2015). Burge on Perception, in E. Margolis and S. Laurence (eds.), *The Conceptual Mind: New Directions in the Study of Concepts. Cambridge*, MA:MIT Press.
- Fougnie, D. 2008: The Relationship between Attention and Working Memory. In N.B.Johansen (Ed.). *New research on short-term memory*. New York: Nova Biomedical Books.
- Frick, R. W. (1984). Using both an auditory and a visual short-term store to increase digit span. *Memory & Cognition*, *12*(5), 507-514.
- Fuller, S., Ling, S., & Carrasco, M. (2004). Attention increases perceived saturation. *Journal of Vision*, *4*(8), 329-329.
- Gauker, C. (2011). Concepts are not icons. *Behavioral and Brain Sciences* 34 (3):127.
- Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animateinanimate distinction as examples. *Cognitive Science*, 14, 79–106
- Gentle, M J, Hunter, L N and Corr, S A 1997 Effects of caudolateral meostriatal ablations or pain related behaviour in the chicken, *Physiology and Behaviour* 61(4):493-498.
- Gibson, J.J. (1977).The Theory of Affordances. In Shaw, R., & Bransford, J. (eds), *Perceiving, acting, and knowing: Toward an ecological psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Godfrey-Smith, P. (2004). On Folk Psychology and Mental Representation.*Representation in Mind*, 147-162.
- Godfrey-Smith, P. (2016). Do squid feel pain? *London Review of Books*, Vol. 38 No. 3, 25-26.
- Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, 6, 298–304.
- Gombrich, E. H. (1983). The Image and the Eye: Further Studies in the Psychology of Pictorial Representation. *Journal of Aesthetics and Art Criticism* 42 (1):85-89.
- Gottfried, J. A. (2011). *Neurobiology of sensation and reward*. Boca Raton, FL: CRC Press.
- Grahek, N. (2008). *Feeling pain and being in pain*. MIT Press, Cambridge, 2nd edition.
- Greene, J. D. (2005). Apraxia, agnosias, and higher visual function abnormalities. *Journal of Neurology, Neurosurgery & Psychiatry, 76*(Supplemental, 5), V25-V34.
- Greene, R. L. (1989) Immediate serial recall of mixed-modality lists. *Journal of Experimental Psychology:*

*Learning, Memory, and Cognition* 15:266–74.

- Grill-Spector, K., & Kanwisher, N. (2005). Visual Recognition. *Psychological Science, 16*(2), 152-160.
- Grosenick, L., Clement, T. S., & Fernald, R. D. (2007). Fish can infer social rank by observation alone. *Nature, 446*(7131), 102.
- Gross, Steven & Flombaum, Jonathan (forthcoming). Does Perceptual Consciousness Overflow Cognitive Access? The Challenge from Probabilistic, Hierarchical Processes. *Mind and Language*.
- Grush, R., 2007: A plug for generic phenomenology. *Behavioral and Brain Sciences*, 305-6.
- Halford, G. S., Phillips, S., & Wilson, W. H. (2001). Processing capacity limits are not explained by storage limits. *Behavioral and Brain Sciences* 24 (1), 123-124.
- Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience Nat Neurosci, 9*(11), 1367-1368.
- Harris, J., Miniussi, C., Harris, I., & Diamond, M. (2002). Transient storage of a tactile memory trace in primary somatosensory cortex. *Journal of Neuroscience*, 22, 8720–8725.
- Hebb, D.O. (1949). *The Organization of Behavior*, Wiley: New York.
- Heck, R. G., 2000. Nonconceptual content and the space of reasons. Philosophical Review, 109: 483–523.
- Herrnstein, R. J., & Loveland, D. H. (1964). Complex Visual Concept in the Pigeon. *Science, 146*(3643), 549-551.
- Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.
- Hölldobler, B., & Wilson, E. O. (1990). *The Ants*. Cambridge, MA: Belknap Press of Harvard University Press.
- Hume, D. *A Treatise of Human Nature*. Oxford, Oxford University Press, 1968.
- Hwang A. D., Wang H.-C., Pomplun M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research* 51, 1192–1205.
- Intraub, H., & Richardson, M. (1989). Wide-angle memories of close-up scenes.*Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(2), 179-187.
- Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105(19), 6829-6833.
- James, W. (1890). *The Principles of Psychology*. Dover Publications.
- James, W. (1900). *Talks to teachers and students*. New York: Henry Holt.
- Jankowiak, J., Kinsbourne, M., Shalev, R. S., & Bachman, D. L. (1992). Preserved Visual Imagery and Categorization in a Case of Associative Visual Agnosia.*Journal of Cognitive Neuroscience, 4*(2), 119-131.
- Jeannerod, M. (2006). *Motor cognition*. Oxford: Oxford University Press.
- Kahane, G. (2009). Pain, Dislike and Experience. *Utilitas*, 21(03), 327-336.
- Karban, R. (2008). Plant behaviour and communication. *Ecology Letters, 11*(7), 727-739.
- King, J., Pescetelli, N., & Dehaene, S. (2016). Selective maintenance mechanisms of seen and unseen sensory features in the human brain.
- Kirkpatrick, K., Bilton, T., Hansen, B. C., & Loschky, L. C. (2014). Scene gist categorization by pigeons. *Journal of Experimental Psychology: Animal Learning and Cognition, 40*(2), 162-177.
- Klein, C. (2015). *What the Body Commands: The Imperative Theory of Pain*. MIT Press.
- Koch, C., & Tsuchiya, N., (2007). Phenomenology without conscious access is a form of consciousness without top-down attention. *Behavioral and Brain Sciences* 30 (5-6):509-510.
- Korsgaard, C. (1996). *The sources of normativity*. Cambridge: Cambridge University Press.

- Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
- Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society* B, 362: 857–875.
- Kouider, S., de Gardelle, V., Sackur, J. and Dupoux, E. (2010). How Rich is Consciousness? The Partial Awareness Hypothesis, *Trends in Cognitive Sciences* 14, pp. 301–307.
- Kouider, S., Dehaene, S., Jobert, A. & Le Bihan, D. (2007). Cerebral bases of subliminal and supraliminal priming during reading. *Cerebral Cortex 17:2019-2029.*
- Kroll, N. E., Parks, T., Parkinson, S.R., Bieber, S.L. & Johnson, A.L. (1970). Short-term memory while shadowing: Recall of visually and of aurally presented letters. *Journal of Experimental Psychology* 85 (2):220.
- LaBuda, C. J., & Fuchs, P. N. (2000). A Behavioral Test Paradigm to Measure the Aversive Quality of Inflammatory and Neuropathic Pain in Rats. *Experimental Neurology* 163: 490–494.
- Lamme, V. A. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*, *1*3, 204-220.
- Lamme, V. A. F. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences* 10 (11):494-501.
- Landman, R., Spekreijse, H. and Lamme, V. (2003). Large Capacity Storage of Integrated Objects Before Change Blindness, *Vision Research* 43:149–164.
- Lau, H. & Brown, R. (forthcoming). The Emperor's New Phenomenology? The Empirical Case for Conscious Experience without First-Order Representations. In Adam Pautz & Daniel Stoljar (eds.), *Themes from Block*. MIT
- Lau, H. (2008). A higher order Bayesian decision theory of consciousness. In Rahul Banerjee & B. K. Chakrabarti (eds.), *Models of Brain and Mind: Physical, Computational, and Psychological Approaches*. Elsevier.
- Levin, D. T., & Banaji, M. R. (2006). Distortions in the perceived lightness of faces: The role of race categories. *Journal of Experimental Psychology: Genera*l, 135, 501–512.
- Lewis, D. (1982) Logic for Equivocators. Nous 16, 431-441. Reprinted in D. Lewis, *Papers in Philosophical Logic*, Oxford, Oxford University Press, pp. 97-110
- Liu, G. T., Ferguson, A. R., Crown, E. D., Bopp, A. C., Miranda, R. C., & Grau, J. W. (2005). Instrumental Learning Within the Rat Spinal Cord: Localization of the Essential Neural Circuit. *Behavioral Neuroscience*, *119*(2), 538-547.
- Logie, R. H., & Cowan, N. (2015). Perspectives on working memory: Introduction to the special issue. *Memory & Cognition*, *43*(3), 315-324.
- Luck, S. J. & Vogel, E. K. (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390:279–81
- Luck, S. J., Vogel, E. K., & Shapiro, K. L. (1996). Word meanings can be accessed but not reported during the attentional blink. *Nature*, *383*(6601), 616-618.
- Lumer, E. D. (1998). Neural Correlates of Perceptual Rivalry in the Human Brain. *Science*, *280*(5371), 1930-1934.
- Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy & Psychology* 6 (4):547-569.
- Lyon, P. (2015). The cognitive cell: Bacterial behavior reconsidered. *Frontiers in Microbiology*, *6*.
- Lyrra, P. (2005). Review of José Luis Bermúdez: Thinking Without Words. *Psyche* 11 (2), March 2005.
- Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature*

*Neuroscience*, *17*(3), 347-356.

- Maguire, J. F., & Howe, P. D. (2016). Failure to detect meaning in RSVP at 27 ms per picture. *Attention, Perception, & Psychophysics* 78(5), 1405-1413.
- Mandelbaum, E. (2013). Thinking is Believing. *Inquiry* 57 (1):55-96.
- Martínez, M. (2011). Imperative content and the painfulness of pain. *Phenom Cogn Sci Phenomenology and the Cognitive Sciences*, 10(1), 67-90.
- Mayer, E. A. (2011). Gut feelings: The emerging biology of gut–brain communication. *Nature Reviews Neuroscience*, *12*(8), 453-466.
- McDowell, J. (1994). *Mind and World*. Cambridge: Harvard University Press.
- Merker, B. (2005). The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition*,14(1), 89-114.
- Mikels, J., Reuter-Lorenz, P., Beyer, J., & Fredrickson, B. (2008). Emotion and working memory: Evidence for domain-specific processes for affective maintenance. *Emotion*, 8, 256–266.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 632, 81-97.
- Mole, C. (2011). *Attention is Cognitive Unison: An Essay in Philosophical Psychology*. Oxford University Press Usa.
- Moores E., Laiti L., Chelazzi L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience* 6, 182–18.
- Morgan, C. L. (1903). *An introduction to comparative psychology*, 2nd edition. London: W. Scott. p. 59.
- Naccache, L., Gaillard, R., Adam, C., Hasboun, D., Clémenceau, S., Baulac, M., Dehaene, S., & Cohen, L. (2005). A direct intracranial record of emotions evoked by subliminal words. *Proceedings of the National Academy of Sciences of the United States of America*. 21. Vol. 102. 2005. pp. 7713–7.
- Nagel, T. 1974. “What is it like to be a bat?” Philosophical Review, 83: 435–456.
- Nakagawa, E. (1993). Matching and non-matching concept learning in rats. *Psychobiology* 21: 142.
- Noë, A. (2005). *Action in Perception*. The MIT Press.
- Nozick, R. (1974). *Anarchy, state, and utopia*. New York: Basic Books.
- O'Regan, J. Kevin & Noë, Alva (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24 (5):883-917.
- Orgs, G., Lange, K., Dombrowski, J., & Heil, M. (2006). Conceptual priming for environmental sounds and words: An ERP study. *Brain and Cognition*, *62*(3), 267-272.
- Papineau, D. (2007). Reuniting (scene) phenomenology with (scene) access. *Behavioral and Brain Sciences* 30 (5-6):521-521.
- Parfit, D. (1984). *Reasons and persons*. Oxford: Clarendon Press.
- Pashler, H. (1989). Dissociations and dependencies between speed and accuracy: Evidence for a two-component theory of divided attention in simple tasks. *Cognitive Psychology*, 21(4), 469-514.
- Pashler, H. (1991). Shifting visual attention and selecting motor responses: Distinct attentional mechanisms. *Journal of Experimental Psychology*: Human Perception and Performance, 17(4), 1023-1040.
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116(2), 220-244
- Perky, C.W. (1910) An Experimental Study of Imagination. *American Journal of Psychology*(21) 422–52.
- Peters, M. A., & Lau, H. (2015). Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *ELife*, 4.

- Peterson, M., & Gibson, B. (1994). Object recognition contributions to figure-ground organization: Operations on outlines and subjective contours. *Perception and Psychophysics* 56(5): 551–564
- Phillips, I. (2015). Consciousness and Criterion: On Block's Case for Unconscious Seeing. *Philosophy and Phenomenological Research* 92 (3).
- Pinto, Y., Sligte, I., & Lamme, V. (2013). Working memory requires focal attention, fragile VSTM does not. *Journal of Vision*, *13*(9), 459-459.
- Potter, M. (2009). Conceptual short term memory. *Scholarpedia*, 52, 3334.
- Potter, M. C. (1993). Very short-term conceptual memory. *Memory & Cognition*, 21, 156-161.
- Potter, M. C. (2012). Conceptual Short Term Memory in Perception and Thought. *Frontiers in Psychology*, 3:113.
- Potter, M. C., Staub, A., & O'Connor, D. H. (2004). Pictorial and conceptual representation of glimpsed pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 478-489.
- Potter, M.C., Kroll, J.F., & Harris, C. (1980). Comprehension and memory in rapid sequential reading. In R. Nickerson Ed., *Attention and Performance VIII* pp. 395-418. Hillsdale, NJ: Erlbaum.
- Potter, M.C., Stiefbold, D., & Moryadas, A. (1998). *Word selection in reading sentences: Preceding versus following contexts. Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 68-100.
- Potter, M.C., Wyble, B., Hagmann, C.E., & McCourt, E.S. (2014): Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, 762, 270-279.
- Premack, D. (1990). The infant's theory of self-propelled motion. *Cognition*, 36, 1–16.
- Prinz, J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. MIT Press.
- Prinz, J. (2005). A Neurofunctional Theory of Consciousness. In A. Brook & K. Akins (eds.), *Cognition and the Brain: The Philosophy and Neuroscience Movement*. Cambridge University Press 381-396.
- Prinz, J. (2007). Accessed, accessible, and inaccessible: Where to draw the phenomenal line. *Behavioral and Brain Sciences*, 305-6.
- Prinz, J. (2011a). The Sensory Basis of Cognitive Phenomenology. In T. Bayne, & M. Montague, (eds.), *Cognitive Phenomenology*. OUP.
- Prinz, J. (2011b). Is Attention Necessary and Sufficient for Consciousness? In C. Mole, D. Smithies & W. Wu (eds.), *Attention: Philosophical and Psychological Essays*. Oxford University Press 174-204.
- Prinz, J. (2012). *The Conscious Brain: How Attention Engenders Experience*. Oxford: Oxford University Press.
- Proffitt, D. R., Stefanucci, J. K., Banton, T., & Epstein, W. (2003). The role of effort in perceiving distance. *Psychological Science*, 14, 106–112.
- Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, *32*(1), 65-97.
- Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(03). doi:10.1017/s0140525x99002022
- Pylyshyn, Z., Burkell, J., Fisher, B., Sears, C., Schmidt, W. & Trick, L. (1994) Multiple parallel access in visual attention. *Canadian Journal of Experimental Psychology* 48:260–83.
- Ramsøy, T. Z. and Overgaard, M. 2004. Introspection and subliminal perception. *Phenomenology and the Cognitive Sciences* 3: 1–23.
- Regan, T., & Singer, P. (1989). *Animal rights and human obligations*. Englewood Cliffs, NJ: Prentice Hall.
- Rosenthal, D. (2005). *Consciousness and Mind*, Oxford: Clarendon Press.
- Rosenthal, D. (2009). Concepts and definitions of consciousness. In P. W. Banks (ed.), *Encyclopedia of*

*Consciousness*. Elsevier.

- Rosenthal, D. (2010). How to think about mental qualities. *Philosophical Issues* 20 (1):368-393.
- Rubens, A. B., & Benson, D. F. (1971). Associative Visual Agnosia. *Archives of Neurology, 24*(4), 305-316.
- Salmela, V. R., Moisala, M., & Alho, K. (2014). Working memory resources are shared across sensory modalities. *Attention, Perception, & Psychophysics*, 767, 1962-1974.
- Saults, J.S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, 1364, 663-684.
- Schwarb, H., Nail, J., & Schumacher, E. H. (2015). Working memory training improves visual short-term memory capacity. *Psychological Research, 80*(1), 128-148.
- Schwitzgebel, E. (2008). The Unreliability of Naive Introspection. *Philosophical Review*, *117*(2), 245-273.
- Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience 8*(10), 1391-1400.
- Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*, *125*(1), 4-27.
- Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology, 22*(2), 261-273.
- Shea, N. (2011). New concepts can be learned. *Biology and Philosophy* 26 (1):129 - 139.
- Shriver, A. (2014). The Asymmetrical Contributions of Pleasure and Pain to Subjective Well-Being. *Review of Philosophy and Psychology Rev.Phil.Psych.,5*(1), 135-153.
- Siegel, S. (2010). *The Contents of Visual Experience*. Oxford.
- Siegel, S. (2014). Affordances and the Contents of Perception. In B. Brogaard (Ed.), *Does Perception Have Content?* Oxford University Press.
- Siewert, C. (2011). Phenomenal thought. In T. Bayne and M. Montague (eds.), *Cognitive Phenomenology*. Oxford University Press.
- Sligte, I., Scholte, H. S. and Lamme, V. (2008). Are There Multiple Visual Short-Term Memory Stores? *PloS ONE* 3, pp. e1699.
- Sloman, S.A. (2002) Two Systems Reasoning. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.) *Heuristics and Biases: The Psychology of Intuitive Judgme*nt. Cambridge University Press.
- Sneddon, L. U., Elwood, R. W., Adamo, S. A., & Leach, M. C. (2014). Defining and assessing animal pain. *Animal Behaviour, 97*, 201-212.
- Sneddon, L. (2013). Do painful sensations and fear exist in fish. In T. A. van der Kemp, & M. Lachance (Eds.), *Animal suffering: From science to law, international symposium*. Toronto, Canada: Carswell, 93-112.
- Sorabji, R. (1993). *Animal Minds and Human Morals: the origins of the Western debate*. Ithaca, NY: Cornell University Press.
- Soto, D., Mäntylä, T., Silvanto, J. (2011). Working memory without consciousness. *Current Biology* 21:R912–3.
- Spearman, C. B. (2005). *The Abilities of Man: Their Nature and Measurement*. The Blackburn Press.
- Sperling, G. (1960). The Information Available in Brief Visual Presentations, *Psychological Monographs: General and Applied* 74, pp. 1–29.
- Stalnaker, R. (1984). *Inquiry*, Cambridge, MA: MIT Press.

- Stalnaker, R. (1998). What might nonconceptual content be? In E. Villanueva (Ed.), *Concepts* (Philosophical Issues, Volume 9), Atascadero: Ridgeview.
- Stazicker, J. (2011). Attention, Visual Consciousness and Indeterminacy. *Mind & Language*, 262, 156-184.
- Stein, T., Kaiser, D., & Hesselmann, G. (2016). Can working memory be non-conscious? *Neurosci Conscious Neuroscience of Consciousness*, *2016*(1).
- Stoerig, P., & Barth, E. (2001). Low-level phenomenal vision despite unilateral destruction of primary visual cortex. *Consciousness & Cognition* 10(4): 574–587.
- Susilo, T., E. McKone, E., & Edwards, M. (2010). Solving the upside-down puzzle: Why do upright and inverted face aftereffects look alike? *Journal of Vision* 10(13): 1–16
- Teuber, H.L. 1968. Alteration of perception and memory in man. In L. Weiskrantz ed. *Analysis of Behavioral Change*. New York: Harper & Row.
- Thompson, P. (1980). Margaret Thatcher: a new illusion. *Perception* 9 (4).
- Tolman E.C. (1948). Cognitive maps in rats and men. *Psychological Review* 55 (4): 189–208
- Tononi, G. (2008). Consciousness as Integrated Information: A Provisional Manifesto. *The Biological Bulletin, 215*(3), 216
- Tononi, G., & Koch, C. (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370* (1668).
- Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing.*Journal of Experimental Psychology: Human Perception and Performance*, *19*(2), 331-351.
- Tsuchiya, N., Wilke, M., Frässle, S., & Lamme, V. A. (2015). No-Report Paradigms: Extracting the True Neural Correlates of Consciousness. *Trends in Cognitive Sciences*, *19*(12), 757-770.
- Tye, M. (1995). *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.
- Ullman, S. (1996). *High-level vision: Object recognition and visual cognition*. Cambridge, MA: MIT Press.
- Vogel, E. K., Luck, S. J. & Shapiro, K. L. (1998) Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance* 24:1656–74.
- Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 92-114.
- Von Wright, J. (1970). On selection in visual immediate memory. *Acta Psychologica*, 33, 280-
- Walker, M. (1980). The *Lexicon of Comicana*. iUniverse.
- Wallace, D. F. (2005). *Consider the lobster and other essays*. New York: Little, Brown.
- Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology* 27: 635–657.
- Weiskrantz, Lawrence (1986). *Blindsight: A Case Study and Implications*. Oxford University Press.
- Witt, J. K., Proffitt, D. R., & Epstein, W. (2004). Perceiving distance: A role of effort and intent. *Perception*, 33, 577–590.
- Wu, W. (2014). *Attention*. Routledge.
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233-235.