Conceptual short-term memory: a missing part of the mind

Abstract

Recent discussion of the role of memory in perception has tended to focus on sensory memory and working memory. I present a summary of the psychological evidence for a further form of memory termed Conceptual Short-Term Memory, arguing it cannot be easily understood in terms of more familiar mechanisms. I suggest that Conceptual Short-Term Memory may shed new light in several philosophical debates, focusing on Ned Block’s overflow argument and the relationship between consciousness and reportability. I conclude with some reflections on the relevance of Conceptual Short-Term Memory for questions about the boundary between perception and cognition.

Henry Shevlin, CUNY Graduate Center

Correspondence address: CUNY Graduate Center, 365 Fifth Avenue, New York. NY10016 US

Email: henry.shevlin@gmail.com
Conceptual short-term memory: a missing part of the mind

Most discussion of the role of memory in perception has focused on the contribution of working memory and sensory memory. However, there is increasingly strong psychological evidence that these do not exhaust the forms of memory at work in perception. In this paper, I describe experimental data that suggests there is a form of short-term high-capacity conceptual memory called Conceptual Short Term Memory (CSTM). I claim that this data cannot be easily understood in terms of other memory mechanisms, and argue that CSTM provides the best explanation for the empirical data presented. Additionally, I claim that CSTM is of potential interest and relevance for philosophy and cognitive science, and may shed new light on several problems in debates about consciousness and perception.

The paper proceeds in three stages. The first part of the paper provides a brief background on the major forms of memory as currently understood, focusing particularly on sensory memory and working memory. The second part of the paper describes the main empirical data that provides evidence for the existence of CSTM. The third part of the paper suggests two areas in philosophy of mind where CSTM may be fruitfully employed, namely the overflow argument of Ned Block and the relationship between consciousness and reportability. The paper concludes with a brief reflection on the question of whether CSTM should be considered a form of perception or cognition.

1. How many stages of memory are involved in perception?

Before spelling out how Conceptual Short-Term Memory can enrich our models of perceptual memory, it is worth outlining the major forms of memory as currently understood. A useful starting point is the historically important and still influential picture of human memory provided by the Atkinson-Shiffrin memory model (Atkinson & Shiffrin, 1968). This model posits that human memory consists of three basic components: 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.

The Atkinson-Shiffrin model has faced considerable criticism. In particular, there is reason to think that more complex mechanisms are operative at each of the levels of processing than initially suggested. For example, short-term memory is now often taken to include both
fragile forms of sensory memory and more robust forms of working memory, where this latter is sometimes understood not just as a simple store, but as an attentionally-gated mechanism for selecting and inhibiting information in various forms of sensory memory (Baddeley and Hitch, 1974). Likewise, long-term memory seems to rely on distinct mechanisms for encoding learned motor skills (procedural memory) and memories of previously learned words and past experiences (semantic and episodic memory), as demonstrated by the preservation of the former but not the latter in some lesion patients (Milner, 1965).

Most modern models of memory, however, can still be understood in broadly tripartite terms, distinguishing between sensory forms of memory, working memory, and long-term memory. What follows will be primarily concerned with memory as operative in moment-to-moment perception, so I will set aside long-term memory, and focus now on describing what is known about sensory memory and working memory.

Consider sensory memory first. 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. For example, George Sperling, in his work on partial report paradigms, famously showed the existence of a form of sensory memory dubbed iconic memory. In his original experiment, subjects were briefly shown a 3x4 matrix of alphanumeric characters (Sperling, 1960). 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.

A further important form of visual sensory memory has been demonstrated much more recently by Sligte and colleagues. This form of memory, called Fragile Visual Short-Term Memory (fVSTM), persists significantly longer than iconic memory, and seems to have an even higher capacity than Sperling’s iconic memory, with an upper limit of at least 32 objects (Sligte et al., 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
that, at least in the case of visual 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 and fVSTM (Sligte, Scholte, & Lamme 2008; Averbach & Coriell, 1961; Saults & Cowan, 2007, show overwriting effects in echoic memory). The second is that while visual sensory memory allows in some cases for feature binding (Landman, R., Spekreijse, H., & Lamme, 2003), it seems to encode objects in respect of their sensory properties such as color and shape, rather than higher-level semantic properties such as conceptual identity. This is illustrated by the discovery that subjects do not exhibit partial report superiority when cued to report just on items belonging to a given semantic category, such as ‘numbers’ or ‘letters’ (Sperling, 1960; von Wright, 1970).

Working memory is quite different from these forms of short-term memory. Whereas sensory memory is rich but fragile, working memory is a robust, flexible, but strictly capacity-limited store recruited for complex attentionally-demanding tasks, such as memorizing lists of unfamiliar numbers or words or picking out a previously seen color patch from an array (Cowan, 2010; Luck & Vogel, 1997).

Early evidence for some capacity-limited form of central short-term memory was presented in the classic paper ‘The Magical Number Seven, Plus or Minus Two’ (Miller, 1956), 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. Additionally, people are frequently able to remember many more than seven items at a time via the use of chunking; that is, combining different words, numbers, or even visual spatial items into a single representation for the purposes of memory storage.

Nonetheless, when experiments are designed to control for chunking (as well as the employment of sensory forms of memory, such as iconic and echoic 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 3-5 items at a time. This holds true regardless of whether the relevant information to be retained is semantic (as in the identity of a digit or the meaning of a word) or tied to a particular sensory modality (as in a color patch or tone). In fact there is evidence that working memory involves a single store for information from different modalities (Salmela, Moisala, & Alho, 2014). Despite the evidence that working memory relies on a single central
store of information, in what follows I will follow convention in using the term visual working memory to describe working memory as operative specifically in the visual domain.¹

The foregoing discussion suggests a complex picture of perceptual processing and memory involving multiple distinct forms of memory. Nonetheless, it is not a gross oversimplification to claim that there are two main forms of memory widely agreed to be involved in perceptual tasks, namely sensory memory and working memory. In the case of vision, the former consists of iconic memory and fragile visual short-term memory, characterized by their high capacity, relatively brief duration, and vulnerability to disruption by new visual inputs, while the latter consists of visual working memory, which is characterized by a relatively small capacity (around four items), its longer duration, and its stability in the face of new visual inputs.

Note that there is neuroscientific evidence from that two forms of memory seem to be localized in different brain regions. For example, iconic memory and fragile visual short-term memory seem to rely on processing in V1 and the extrastriate cortex, and can be rendered inoperative using transcranial magnetic stimulation to V1 (Jolij & Lamme, 2005). Conversely, visual working memory seems to rely on the dorsolateral prefrontal cortex: selective transcranial magnetic stimulation of this region significantly impairs performance on visual working memory tasks while leaving fragile visual short-term memory intact (Sligte, Wokke, Tesselaar, Scholte, & Lamme, 2011).

Before proceeding, it is worth noting that there is ongoing debate concerning how these forms of memory relate to attention. It is sometimes assumed that sensory forms of memory are preattentive (Strous, Grochowski, Cowan, & Javitt, 1995; Lamme, 2003). Recent work, however, provides some evidence that maintaining and accessing iconic memory in fact requires attention (Persuh, Genzer, & Melara, 2012). There is also little consensus on the relationship between working memory and attention. While updating the contents of working memory seems to be an attentionally demanding process, there is also good reason to think that at least some attentionally-demanding tasks do not interfere with maintenance of information in working memory, depending on the sensory modalities and types of task involved (Fougnie, 2008). I will

¹ I will also assume for the purposes of this paper that working memory is always conscious. However, this is a subject of recent and ongoing controversy; see Stein, Kaiser, and Hesselmann (2016) for a review of the debate.
therefore aim to avoid bringing attention into the discussion that follows save where absolutely necessary.

2. Conceptual short-term memory: a new form of perceptual memory?

I now turn to the major goal of this paper, which is to summarize recent evidence suggesting the existence of a further form of memory involved in perceptual processing, namely conceptual short-term memory (CSTM). This is a proposed form of short-term memory that is not identical either with sensory forms of memory or the higher level mechanisms of visual working memory. CSTM was initially posited by Mary Potter, who 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, 2010). Unlike sensory memory, it encodes high-level features of objects and scenes and is not vulnerable to disruption by new visual inputs. It also has a shorter duration than some of the more robust forms of sensory memory such as fragile visual short-term memory. However, unlike visual working memory, it has a high capacity and has a very brief duration.

I will now examine some of Potter’s evidence for CSTM in more detail, before considering some debunking explanations. In a research program spanning several decades, Potter demonstrated that subjects rapidly classify sequences of rapidly presented images and words and are able to selectively access parts of this semantic information, with unaccessed information being almost immediately lost. For example, subjects rapidly and sequentially presented with the words from a twelve-word sentence can accurately remember every word they see, but cannot do so for a similarly presented list of twelve unrelated words (Potter, Kroll, & Harris, 1980). This suggests that subjects must be 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 working memory via chunking mechanisms, but in cases where those connections do not exist, the information is almost immediately lost.

Another experiment of Potter’s provided further evidence that this conceptual information is rapidly lost unless consolidated (see Fig. 1). 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 they had rejected, even though they presumably must have understood its meaning in order to realize that it was the less contextually appropriate of the two words (Potter, Stiefbold, & Moryadas, 1998).

<Figure 1 here>

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 working memory. More recent work of Potter’s is harder to explain away in terms of working memory, however (Potter, Wyble, Hagmann, & McCourt, 2014). In one experiment, subjects were shown six or twelve sequential images (which they had not previously seen) for durations of 13, 27, 53, or 80ms (see Fig. 2). They were given a target description (for example, ‘a 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 longer the initial duration of the stimuli, the more likely subjects were to correctly detect a target, but were above chance in all trials. However, 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 (with the exception of the last picture, which they were not tested on).

<Figure 2 here>

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 presented 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 uncontroversial. What is much more open for debate is which memory mechanisms underpin subjects’ performance. Potter’s own interpretation, and my own, is that this result is best explained in terms of CSTM. Specifically, I believe 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) and 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 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 debunking approaches also are worth considering. Specifically, it is not immediately obvious that the results cannot be explained just in terms of iconic memory or fragile visual short-term memory, or perhaps even ordinary visual working memory.

Consider first the hypothesis that subjects’ performance might be explained in terms of ordinary visual working memory. This might 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 access semantic information about each image as it was 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 earlier, visual working memory has strict capacity limits, well below the twelve images presented in some trials. At most, then,
subjects could retain 3-4 images in visual 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 visual working memory.\(^2\)

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 in the Landman experiment. 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 rapidly conceptualize and ‘inspect’ if and when they are given a target description.

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 and Landman trials, since these involve just a single initial stimulus prior to cueing. In this experiment, however, subjects saw multiple images one after another in exactly the same location prior to cueing. Any representations in iconic or fragile visual short-term memory would therefore be rapidly ‘overwritten’ as the sequence was presented.\(^3\)

Additionally, such an interpretation would struggle to explain similar rapid serial visual presentation experiments performed by Potter that examined subjects’ vulnerability to conceptual ‘decoys’. In one such experiment, subjects were shown a sequence of five pictures for 173ms, 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.

---

\(^2\) The authors commented that “we can reject the hypothesis that participants could encode only two or three pictures in working memory; 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).

\(^3\) One complicating factor comes from the fact that Potter has elsewhere suggested that there may be another new form of very brief duration ‘pictorial short-term memory’ (PTSM) that serves as a short-term sensory store which is not vulnerable to overwriting by serial presentation. This is posited by the authors to explain subjects’ improved performance in very short test intervals at visually recognizing pictures under forced choice relative to identifying them under a given description. Little further work has been done on PTSM, however, and in any case, it cannot explain most of the effects attributed to CSTM. Note also that Potter et al. consider it possible that the performance difference they observed might also be explained purely in terms of the more rapid loss of fine-grained relative to coarse-grained conceptual information (see Potter, Staub, & O’Connor, 2004:487-8).
However, they were significantly more prone to error when tested on distractors which were similar in 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. 3).

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 strong evidence that subjects’ performance in Potter’s work on CSTM is not simply a matter of sensory short-term memory.

<Figure 3 here>

Potter’s work provides good reason for positing a stage of processing like CSTM. However, I now wish to present a quite different form of empirical evidence for the CSTM hypothesis which comes from recent work on visual search tasks. Unlike Potter’s experiments, most of which involve rapid serial visual presentation, the experiments I’ll now describe involves presentation of a single array of stimuli, much like the Sperling and Landman studies discussed earlier.

The first such experiment involved the presentation of a single grid of objects to subjects for brief durations (Moores, Laiti, & Chelazzi, 2003). 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. Moores showed that the presence of semantically-related distractors in the array (e.g., a motorbike helmet) had a negative impact on subjects’ reaction times and accuracy in trials where the target was absent. This suggests that subjects were diverted by the presence of pictures that were conceptually similar 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 scanned the array they 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 the semantic properties of stimuli (subjects in the Sperling task could not be trained to report just on letters or just on numbers, for example). However, in this trial, subjects’ eye movements were sensitive to semantic properties of items in the array.

<Figure 4 here>

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.

This interpretation looks much less plausible, however, in light of a later similar experiment (Belke, Humphreys, Watson, Meyer, & Telling, 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 not limited by the resources of working memory.4

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. 5). If subjects were using capacity-limited working memory mechanisms to

---

4 It remains unclear whether there is still a role for some form of attention in the rapid retrieval of semantic information seen in this experiment. It may be significant that subjects are explicitly instructed to identify objects fitting a given description, thus making semantically-related items relevant to their task and thus a target for attention. By contrast, in paradigms such as inattentional blindness (such as Mack & Rock, 1998), subjects are typically engaged in a cognitively demanding task while a task-irrelevant item is presented in the background.
rapidly identify the semantic content of each item in the array, one would expect their performance to be severely hindered by this task.

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 the initial looking behavior of subjects at all; subjects 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 (for the purposes of directing initial eye movements) does not rely on the same cognitive resources as working memory. 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).

To summarize, then, we have rich data from multiple paradigms suggesting that there is short-term semantic processing of visually presented information outside of capacity-limited working memory which does not vulnerable to disruption by new visual input. The conceptual short-term memory 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 the CSTM seriously.

3. CSTM and consciousness

If this CSTM hypothesis is correct, then it is of considerable interest for anyone studying the mechanisms of human perception. Potter herself, for example, believes CSTM is required “to

---

5 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 in 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.
explain the human ability to understand and act rapidly, accurately, and seemingly effortlessly in response to the presentation of richly structured sensory input.’

In the next part of this paper, however, I wish briefly to examine the significance of the CSTM hypothesis for two more theoretical debates concerning consciousness, namely Block’s overflow argument, and the question of whether conscious states must be reportable. In neither case will I attempt to give a worked-out theory, but I will suggest ways in which CSTM might be shed some new light on the debates, and provide for fruitful new directions for theorists of consciousness.

Before proceeding, however, I should mention one major challenge to applying CSTM in this way, which is that it is very much an open question whether representations in CSTM are ever themselves conscious. Subjects’ reports alone are unlikely to settle the matter, since what is actually reported by subjects will always be limited by what can be encoded in working memory.

However, I think the hypothesis that CSTM is conscious is certainly worth exploring. Additionally, a defender of such a view can appeal to the same kinds of theoretical considerations as theorists like Ned Block and Victor Lamme who claim that forms of sensory memory are conscious (Block 2007, Lamme 2010). As I will go on to argue, in the case of both CSTM and iconic memory, psychology provides us with independent evidence for short-term memory stores which could, if they are conscious, explain certain experimental subjects’ judgments about their experience, as well as some broader phenomenological data. Just like the case for consciousness in iconic memory, then, the case CSTM’s being conscious will rest on an inference to the best explanation, and will draw on broader theoretical and phenomenological considerations as well as future psychological and neuroscientific results.

3.1 - CSTM and overflow

I now turn to the two debates in question. Many readers will doubtless be familiar with Ned Block’s famous claim that ‘phenomenology overflows cognitive accessibility’ (Block, 2007: 494). The central idea in Block’s argument is that at any given time, a subject may be phenomenally conscious of more items than they can cognitively access; that is, what it is like for a subject at any given moment is not just a matter of what they actually notice, think about, or

---

6 Potter herself considers it inconclusive whether CSTM is conscious, 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)
attend to (where these processes are understood to be constrained by working memory).

Block’s argument is developed over numerous articles and appeals both to phenomenological considerations and to empirical work. The basic phenomenological appeal of the overflow claim stems from cases in which it seems to many people to be plausible that the perceptual contents we access in working memory constitute only a subset of the overall phenomenally conscious contents of our experience. Consider, for example, the noise produced by an air conditioner running in the background: most of us are not consciously thinking about it at any given time, yet nonetheless it may seem plausible that the noise contributes to the overall phenomenal character of your experience, as evinced by the feeling, upon noticing it, that one had been hearing it all along.\footnote{The phenomenological appeal of the overflow claim has come in for empirical investigation in its own right. See Schwitzgebel 2007 for an attempted experimental assessment of people’s intuitions about the richness of their experience.} Another kind of example comes from contexts in which we are overwhelmed by sensory information. Thus imagine being in Times Square, surrounded by noise and lights and people and smells. In contexts such as this, it may seem as if we can only cognitively access a small portion of the contents of our experience at any one moment, while also having the impression of consciously seeing and hearing far more than that.

Both kinds of case arguably provide some introspective reason for thinking that phenomenology overflows cognitive access. However, Block also appeals to the work of Sperling and Landman, on iconic memory and fragile visual short-term memory respectively, as providing evidence for overflow. Specifically, Block notes that in several experiments demonstrating the high capacity of sensory memory, subjects are under the impression of having really 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).

Roughly, then, 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 sensory memory is (at least sometimes) conscious even when it not cognitively accessed.

Block’s arguments for overflow have faced a variety of sophisticated and detailed replies, too numerous and too varied in their details to discussed at length here. Many of the proposals have attempted 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).

In contrast to these replies to Block, which deny overflow outright, the CSTM hypothesis provides some options for taking overflow at face value while avoiding certain controversial theoretical commitments bound up in Block’s own position. To see how this might work, note that, like iconic memory, CSTM seems to be operative in scenes involving complex visual arrays (Belke et al., 2008). Moreover, CSTM has a much larger capacity than working memory, and is very fragile, such that only a small capacity of what is stored is able to be accessed for report before the remainder degrades.

Thus if we are looking for a high-capacity fragile memory store that might serve as the constitutive basis for phenomenology without access, CSTM seems as much a suitable candidate as Block’s own candidate, iconic memory, and on the hypothesis that the rich representations in CSTM are conscious, we might capture all the same phenomenological intuitions as Block, as well as allowing us to take at face value subjects’ conviction in the Sperling and Landman trials that they really saw all of the presented items in specific detail.

It is worth emphasizing that this approach would claim specifically that CSTM explained subjects’ reports in the Sperling and Landman trials, not their performance. Subjects’ performance on these trials is clearly dependent on sensory memory, as shown by the fact that partial report superiority is erased by a pattern mask. However, we can appeal to the rich representations in CSTM to explain subjects’ reports about their experience while also claiming the information in sensory memory itself is unconscious until accessed by other mechanisms such as working memory.

---

8A quite different sort of reply is given 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.
Why might this route be preferable to Block’s own hypothesis? There are at least two reasons that this conceptual version of overflow might have particular appeal. First, note that Block’s position involves the idea that non-conceptual states (such as the sensory representations in iconic memory) can be conscious in their own right, independent of whether they give rise to conceptualized states downstream in later processing (Block, 2014a). Thus many – if not most – of our conscious states will on this view consist of iconic (or ‘pictorial’) representations that are not conceptualized by a subject in any way. For theorists who are committed to the idea that the contents of experience are exclusively conceptual (such as McDowell, 1994, and Brewer 1999), this will rule out Block’s account of overflow from the start.

By contrast, nonconceptual content need not play any role in the CSTM account of overflow.9 Hence any theorist who was antecedently sympathetic to the phenomenological arguments in support of overflow but wished to remain a Conceptualist about the contents of experience can appeal to CSTM to reconcile these otherwise competing considerations.

A second reason one might prefer a CSTM account of overflow to Block’s own sensory account concerns high-level perceptual phenomenology. Susanna Siegel and Tim Bayne have recently argued that the phenomenology of experience is not just limited to low-level features such as color and shape but represents a range of high-level features such as natural kinds, individuals, and causal relations (Siegel 2010, Bayne 2009). Thus Siegel suggests 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. These high-level phenomenal properties, Siegel claims, are not just novel concatenations of lower-level color and shape properties, but instead are new forms of phenomenology that develop as one becomes familiar with different kinds of objects and individuals.

If one is sympathetic both to overflow and to Siegel’s claim that visual experiences include such high-level phenomenology, then one may also be likely to think that high-level perceptual phenomenology can occur in overflow cases, that is, among those conscious representations that are not accessed by working memory.10 Although Siegel does not explicitly

---

9 A CSTM account of perceptual experience such as the one sketched here could nonetheless accommodate nonconceptual content. For example, one might claim that perception involves complex representations consisting of both conceptual and nonconceptual content; Carruthers (2015) defends a view along these lines.

10 One might instead claim that high-level phenomenology is a strictly cognitive phenomenon, only arising after perceptual representations are accessed by working memory. However, as discussed below, this is not Siegel’s view.
mention overflow, at times she seems to have such a position in mind, as in the following passage arguing against the idea that high-level phenomenology is a purely cognitive phenomenon.

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 I 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 high-level properties that happens effortlessly and automatically and the more reflective cognitive awareness of such properties that comes when we notice or pay attention to things. However, it is hard to see how to make sense of such a contrast if, like Block or Lamme, one claims that iconic memory is the constitutive basis for conscious perception outside working memory. After all, iconic memory does not encode the high-level semantic properties of scenes that here concern Siegel, nor is there any evidence that a nonconceptual memory store could do so.

Thus in order to defend Siegel’s proposed distinction between the fleeting perceptual representation of high-level properties and their more robust presence in cognition (where this latter presumably involves working memory), one must appeal to some form of phenomenally conscious short-term mechanism that both encodes objects in respect of their semantic properties and yet is distinct from working memory. It may yet emerge that there is some nonconceptual short-term memory store that encodes objects in respect of these semantic properties. As matters stand, however, I know of no candidate empirical mechanism for this aside from CSTM. This provides another reason why one may wish to invoke CSTM in explanations of overflow.

3.2 - Consciousness and reportability

The examples of overflow considered in the previous section concern cases in which experimental subjects believe they have seen more than they can report, as well as instances where our phenomenological judgments suggest that we may be conscious of more than we
access in working memory. However, there is a closely related debate concerning whether we sometimes consciously experience things while believing that we have not experienced them at all.\textsuperscript{11} Several theories of conscious experience claim that whenever we consciously perceive something, we thereby \textit{know} that we consciously perceived it, and can thus report it (and likewise, if we sincerely deny having seen something, then we did not consciously see it).

There are a variety of motivations for this view. One is simply the fact that as a practical matter, psychological experiments depend a great deal on subjects’ report as the decisive measure of consciousness. By contrast, \textit{objective} measures that look for behavioral responses provide evidence that a stimulus has been perceived but do not tell between conscious and unconscious forms of perception. In psychological work, then, report is closely bound up with attributions of conscious perception.

While this might lead one to adopt, for practical purposes, the position that report is the only clear or decisive evidence that a state is conscious, it does not by itself rule out the possibility of conscious states that a subject cannot report (Block, 2007). To make this further claim requires additional conceptual or theoretical work. Dehaene and Naccache, for example, offer a version of global workspace theory that claims that consciousness consists in a form of cognitive access that automatically results in a state’s becoming available for report (Dehaene & Naccache, 2001; Dehaene, 2014). Some higher-order thought theorists of consciousness have also defended a close relationship between consciousness and report, arguing, first, that conscious states are states we are aware of, and second, that the most plausible implementation of this awareness is via higher-order thoughts. These higher-order thoughts then suffice for the relevant first-order state being reportable, such that negative reports indicate that the first-order state is not conscious (Rosenthal, 2005).\textsuperscript{12}

Whatever the merits of these theories, there is nonetheless some phenomenological motivation for resisting the conclusion that we must always be able to report the contents of our

\textsuperscript{11} The debates about overflow and reportability are closely related, but some differences should be kept in mind. Note, for example, that most of the putative cases of overflow (including subjects’ reports in partial report paradigms) are instances where people believe they consciously perceived more than they cognitively accessed. This should be contrasted with cases where subjects confidently report having perceived nothing at all. One could allow that the former but not the latter sort of case can involve conscious experience, on the grounds, for example, subjects must always be in some way cognizant of their conscious states (that is, they must have “access to phenomenality”; see Balog, 2007).

\textsuperscript{12} This is a highly condensed summary that is not intended to capture the nuances of the higher-order approach to consciousness; see Rosenthal & Weisburg (2008) for a more detailed discussion of higher-order theories and their relation to reportability constraints on consciousness.
conscious experiences. One such motivation comes from ordinary cases of visual experiences in which we apparently experience things without realizing we have done so. For example, imagine that I am asked to look at a multicolored object like a Buddhist mandala. Having scanned it, I am immediately asked whether I saw any red patches in the image, and I claim that I didn’t. Upon repeating the task however, I realize that there were indeed a couple of red patches which I had looked at but not noticed. In cases like this, it seems at least an open possibility that we underwent brief conscious experiences that were not reflected in our reports. Likewise, it seems as though we sometimes retroactively correct our reports about the character of our conscious experience; we sometimes mistake an initial sensation of very cold water as being hot, for example, or mistake the feeling of vibration from a dentist’s drill for a feeling of pain. These cases all seem difficult to square with a strict thesis about the relationship between consciousness and reportability.

The vexed question of the relationship between consciousness and reportability has also been highlighted by inattentional blindness paradigms (Mack & Rock, 1998; Simons & Chabris, 1999). These typically involve assigning attentionally demanding tasks during which some striking and bizarre stimulus (such as a man in a gorilla suit) appears in the background. Subjects are typically very surprised to discover (after the fact) that the stimulus was in fact present. A more commonplace case of inattentional blindness might involve a failure to notice a cyclist while driving and being distracted by talking on a mobile phone. But it is controversial how we should interpret such cases: do subjects genuinely not experience the unnoticed items, or do they experience in a way that does not allow for reporting their specific identities?

The question of how to interpret cases like the above is controversial, but what is tacitly assumed in many of these debates is that if there is any notion of conscious experience without report to be had, then the relevant content of that experience must be nonconceptual. Thus many major opponents of the claim that consciousness entails reportability suggest that these cases all

---

13 Rosenthal (2005: 127) concludes that such patients’ experience in fact had the subjective character of pain as a result of misrepresentation at the higher-order level.
14 There are more subtle distinctions here than I have had been able to explore in the main text. For example, one might distinguish between wholly unreportable states (such as, perhaps, those found in conditions like visuospatial neglect), and states that typically fail to be reported because they are almost immediately forgotten. The CSTM view sketched below would allow for consciousness in the latter case, but not necessarily in the former.
15 See Wu (2014) for a summary of the numerous viable interpretations of inattentional blindness.
involve sensory nonconceptual states (Block, 2007; Dretske, 1981; Lamme, 2010).\footnote{Block (2014b) adopts a similar assumption in his discussion of adaptation phenomena. He considers experiments in which subjects claim not to have seen a grid as oriented, but exhibit adaptation effects suggesting that did in fact see it as oriented. He assumes that this must be a nonconceptual effect on the grounds that subjects do not report having seen the grids as oriented. However, if CSTM is not always reportable, it is possible that subjects’ adaptation had a conceptual component without being reportable.}

For Block and Lamme at least, one of the major motivations for this view is perhaps that they assume that any conscious conceptualized states would have to be stored in working memory and hence would be reportable. However, the CSTM hypothesis challenges this latter assumption. Recall that information in CSTM decays extremely rapidly, and subjects in Potters’ experiments were typically unable to recall presented pictures or words if cued more than a second or two after presentation. Note also that noticing an object for the purposes of report or voluntary action (as opposed to merely encoding semantic information about it) requires the involvement of working memory. Hence if CSTM is sometimes conscious, then there may be conscious conceptual representations in CSTM that subjects fail to notice and therefore might deny having seen. Thus CSTM can allow for failures of reporting for conscious states without invoking the notion of nonconceptual content.

What might be the advantages of this? As in the case of overflow, a CSTM account creates new options for a theorist who is sympathetic to Conceptualism about the contents of experience but reluctant to endorse a strict thesis about the relation between consciousness and reportability, specifically allowing them to give an empirically-grounded account of how someone might have conscious experiences that they nonetheless denied having.

However, an arguably even more significant advantage for this approach comes from the constraints it places on the range of states that are candidates for consciousness. 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, allowing, for example, for the possibility of conscious sensory states that the subject does not merely deny having but which are not cognitively accessible to her in any way at all.\footnote{Lamme (2010) allows for inaccessible states, and this seems to be position defended in Block (2007). Block seems to have distanced himself from this possibility in more recent work (2014a: 174-5), instead stressing a more significant link between consciousness and cognitive accessibility.} This creates the threat of what Block terms ‘panpsychic disaster’, in which any we are forced to treat any sort of neural activity at all as a potential realizer of conscious states, even activations in simple parts of the nervous system (such as the LGN or cerebellum) that normally taken to be entirely subpersonal.
A view of perceptual experience that takes encoding in CSTM to be a necessary condition of consciousness, however, can place tighter constraints on the kind of states that might be conscious while simultaneously avoiding committing to a strong reportability claim. The view can claim that all conscious states must be conceptualized in some way or another (including, potentially, in terms of acquired high-level concepts like ‘wedding’ or ‘picnic’) while holding that these states are only fleetingly available to working memory and may pass unnoticed. By requiring that any conscious state be conceptualized by the subject, such a theory presumes a greater degree of connectivity between sensory areas and a subject’s broader mental economy than a model that allows for consciousness wholly independent of cognition, thus ruling out many of the more outlandish examples raised against Block and Lamme, such as wholly inaccessible conscious states in early vision or other presumably subpersonal systems.¹⁸

4. Is this perception or cognition?

My central goal in this paper has been to present evidence that CSTM is a distinctive form of perceptual memory, and as such should be incorporated into theories of perception and memory. I have also attempted to suggest that by adopting the working hypothesis that contents in CSTM can be conscious, we might open up new theoretical options in debates about consciousness, specifically debates about overflow and the relationship between consciousness and reportability.

One lingering question that may be on readers’ minds, however, is whether CSTM should be considered a form of perception or cognition. I have referred to CSTM somewhat loosely as a form of perceptual memory, but it exhibits a strange mix of the usual features that distinguish perception from cognition. For example, while it operates extremely rapidly and seemingly involuntarily, and is not automatically available for report, it nonetheless encodes information in a conceptual rather than purely sensory format. This means that it does easily fit into existing models of perception and cognition. While there is no universal consensus on how to distinguish perception from cognition, one major way of carving up the two (Dretske, 1981) claims that perception is distinguished by its nonconceptual content, while cognition involves propositional

---

¹⁸ Another way to rule out such states is to insist that only sensory states that are accessible to a subject’s cognitive systems can be conscious. This sort of view is held by Prinz (2012) and Carruthers (2011). However, these theories face the challenge of explaining how a state’s merely being disposed to be accessed could be responsible for the occurrent phenomenon of consciousness. A CSTM view, by contrast, can appeal to an occurrent process, namely active encoding in CSTM, to explain why these states are consciously experienced.
attitudes such as believing, judging, and thinking. But while the conceptual format of CSTM might seem to place it on the cognitive side of this divide, it does not seem to support any of our ordinary propositional attitude ascriptions: there is no reason to think that a subject who conceptually classifies an image in CSTM as being a picture of a wedding thereby *thinks* or *believes* that the picture is of a wedding, for example.

Potter herself has claimed that CSTM shows that ‘perception is continuous with cognition’ (Potter, 2012:9). This may be one way to interpret her results, but another would be to maintain that perception and cognition are distinct, and claim that CSTM is different enough from both that it should be considered a further state of processing in its own right.\(^\text{19}\) The results discussed in this paper should serve as a reminder that our pretheoretical intuitions about how best to carve up the mind may not survive the discoveries of psychology and neurosciences fully intact, and that we should be open to possible revisions of our picture of the mind in accordance with new scientific evidence.

Henry Shevlin

Philosophy Department

CUNY Graduate Center

---

\(^\text{19}\) If we really should think of CSTM as indeed distinct from both perception and cognition, then the history of philosophy and psychology can perhaps furnish us with an apt name for it in the notion of “apperception”. As James (1900) puts it, “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.”
REFERENCES


Schwitzgebel, E. 2007: Do you have constant tactile experience of your feet in your shoes? Or is experience limited to what's in attention? *Journal of Consciousness Studies* 14 3:5-35.


