sciendo

Can the Unconscious Image Save "No Overflow"?

Nicholas D'Aloisio-Montilla
University of Oxford

DOI: 10.2478/disp-2018-0002 BIBLID [0873-626X (2018) 48; pp.1–42]

Abstract
The question of whether phenomenal consciousness is limited to the capacity of cognitive access remains a contentious issue in philosophy. Overflow theorists argue that the capacity of conscious experience outstrips the capacity of cognitive access. This paper demonstrates a resolution to the overflow debate is found in acknowledging a difference in phenomenological timing required by both sides. It makes clear that the "no overflow" view requires subjects to, at the bare minimum, generate an unconscious visual image of previously presented items if it is to explain performance in the change detection paradigm. It then demonstrates that conscious imagery should support better task performance than unconscious imagery because of a necessary difference in representational strength. However, this contradicts empirical findings, and so a new argument for overflow is presented without requiring the premise that subjects need to obtain a specific phenomenology of presented items during change detection.

Keywords
Consciousness, imagery, overflow, unconscious, access.

1 The overflow debate

Phenomenal consciousness is what an experience is *like* (Block 1995). A mental state is phenomenally conscious if and only if there is a first-person experience tied to that state—there must be *something that it is like* to undergo a conscious, mental state (Nagel 1974). By contrast, cognitive access is what is available to a subject's cognitive processing (Block 1995, 2007, 2008, 2011). More precisely, information is cognitively accessed if and only if it is available to the processing of a range of consumer subsystems, such as those systems that make use

Disputatio, Vol. X, No. 48, May 2018
Received: 18/02/2018 Revised: 02/10/2018 Accepted: 22/10/2018

of the contents of working memory or the global workspace (Block 2008, 2007; Baars 2002, 2013; Dehaene *et al.* 2006, 1998; Dehaene 2014; Carruthers 2015; Prinz 2012). As a result, accessed information is "actually received by consuming systems", and must be available to subsequent processing without requiring further attentional amplification (Carruthers 2015: 1). For instance, workspace models of consciousness have suggested that cognitive access occurs when information is "globally broadcasted" throughout a functional workspace (Baars 1988; Dehaene *et al*. 2006).

The question of whether phenomenal consciousness and cognitive access are realized by distinct neural machineries remains a key controversy in philosophy (Phillips 2016; Gross and Flombaum 2017; Stacizker 2011; D'Aloisio-Montilla 2017, 2018a, 2018b; Block 2011). To support this disassociation, the "overflow" argument appeals to the recorded capacity of iconic memory to argue that consciousness has a higher capacity than cognitive access: when observing a detailed scene, it is claimed that subjects are capable of consciously representing more than they can bring to report or direct rational control (Block 2007, 2008, 2011, 2014). The primary illustration of such a capacity difference lies in Sperling's paradigm (Sperling 1960; see Phillips 2011a, 2011b for exposition). In this paradigm, a visual stimulus is momentarily presented to subjects. The stimulus is typically a 3x4 grid of letters, and it is maintained for no more than a few hundred milliseconds. When subjects are asked to freely recall the letters of the grid following its disappearance, they correctly remember just 4 or 5 letters. However, Sperling (1960) demonstrated that when subjects are *cued* to a particular row up to 50ms after the grid's offset, then they manage to report 3 or 4 letters from the cued row (Block 2011, Vandenbroucke *et al*. 2011). The "post-stimulus" cue therefore enables subjects to correctly recall the identities of 4 letters from *any one of three rows* in the originally presented grid.

As such, the increased performance ability that results from cueing subjects to a specific subset of letters has been labelled the "partial report advantage" (Sperling 1960, Phillips 2011b). Since cues may direct subject's attention to any row of the grid, *all or almost all* 12 letters of the grid must be encoded into iconic representations while the stimulus is still presently available (Sperling 1960, Block 2011, Phillips 2011a, cf. Gross and Flombaum 2017). Importantly, subjects claim that are able to consciously "see" all 12 letters of the stimulus prior to the cue's onset, even if their reporting abilities are limited to the contents of one row (Block 2008, 2011; Bronfman *et al*. 2014; D'Aloisio-Montilla 2017). In fact, subjects insist that their experience is detailed enough that the conscious content includes the *specific identities* of almost all 12 letters of the grid. Since readout from memory is a necessary marker for access—and given that subjects do not recall the identities of uncued letters—overflow theorists insist that subjects are conscious of more than they access during the task. Thus, phenomenal consciousness appears to "overflow" the capacity of access (Block 2011, 1995).

Overflow opponents argue that subjects *overestimate* their capacity for phenomenology during Sperling's paradigm. The "no-overflow" account maintains that consciousness minimally requires the set of attentional processes that are sufficient for cognitive access (Cohen and Dennett 2011; Phillips 2011a, 2011b; Cohen *et al*. 2012; Stazicker 2011; Kouider *et al*. 2010; Brown 2014; Philips 2016; Gross and Flombaum 2017; Dehaene *et al*. 2006; cf. Block 2011, cf. Landman *et al*. 2003). No-overflow theorists argue that subjects can only form a *specific*[1] conscious experience of the grid's letters *after* top-down attention has been appropriately directed to iconic representations on the presentation of the post-stimulus cue. There are diverging views on which attentional processes are necessary for content to be rendered specifically conscious on this view, with contemporary candidates including working memory encoding (Carruthers 2015), availability to working memory encoding (Prinz 2012), higher order representation (Brown 2014, Lau and Rosenthal 2011), and entrance into the global workspace (Baars 1988, Dehaene *et al*. 2006). Nonetheless, since all candidate theories accept that attention is necessarily *limited* in capacity, subjects are said to only form a *vague impression* of most of the grid's letters during Sperling's paradigm[2]. This is be-

[1] "Specific" in this context entails that conscious experience includes the *identities* of letters.

[2] An alternative interpretation of the no-overflow position is offered by Carruthers (2015). He suggests that subjects can form an *identity-specific* experience of all 12 letters while it is visible to subjects. For an argument against this, see Section 6 of D'Aloisio-Montilla 2017.

cause there are not enough attentional resources to allocate to uncued letters so that they can also receive the necessary amplification required for conscious access. Thus, no-overflow theorists maintain that subjects produce a conscious experience of uncued letters that is, at best,[3] "generic" or "gist-like", and it is this *impression* which they wrongly introspect as a specific conscious experience of almost all letters (Kouider *et al*. 2010, Brown 2014, Lau and Rosenthal 2011).

For a partial report advantage to occur, cues must be presented to subjects strictly *after* the grid is removed from subjects' view. It is therefore claimed that the post-stimulus cue is responsible for occasioning the attentional amplification required for consciousness. Together, these facts imply that on the no-overflow account, a conscious experience of the reported letters must be generated after initial perception of the grid has *ceased* (i.e. after the stimulus is offset). Thus, in contrast with the overflow view, which allows subjects to perceive reported letters with a conscious percept (or "icon") while the stimulus is *still visible*, on the no-overflow view, subjects must *first* summon a specific phenomenology of the cued letters with a "nonperceptual image" or "visual memory image" after the stimulus has been removed (Phillips 2011a: 210; Block 2011; Phillips 2011b). Hence, it seems that the no-overflow account could implicate a mode of highly atypical[4] "visual imagery"—that is, the construction of visual representations that occurs in the absence of external stimuli or relevant perceptual activity[5]—in order to explain how subjects conjure a delayed phenomenology of the cued letters (Kosslyn 2006; Brockmole *et al*. 2002; Phillips 2011a, 2011b, 2014; Nanay 2013; D'Aloisio-Montilla 2017, 2018a).

This paper will provide a new argument for overflow by drawing on the performance of subjects in a recently updated version of

---

[3] Some theorists argue that unattended letters can remain entirely *unconscious* during the task. That is to say that no phenomenology—gist-like, or otherwise—is formed of unattended letters (Cohen and Dennett 2011).

[4] The necessary atypicality of any image employed arises from the timescale in Sperling's task: mental images are typically thought of taking 1–1.5 seconds to generate, and in Sperling's task, subjects form a conscious image of the cued item almost immediately after the grid's offset (Kosslyn 2006).

[5] A more precise definition of visual imagery is presented in Sections 2 and 3.

Sperling's paradigm known as "change detection". This paradigm is useful to the debate because no-overflow theorists must account for the construction of a phenomenology that *first* arises in a significantly more delayed timescale than in the Sperling case. In doing so, the employment of *typical* visual imagery—that is, images which take at least *1–1.5 seconds* to be constructed—can be implicated because cues occur at least 1–4 seconds after perceptual offset unlike in the Sperling case (Kosslyn 2006; D'Aloisio-Montilla 2017, 2018a; Phillips 2011a, 2011b; Block 2011). In change detection, subjects are first presented with a memory array containing eight or so items. Subjects are then presented with a probe array. The items in the probe array are identical to the those in the memory array except one item might be changed. In the blank interval between the memory array's offset and the probe array's onset, subjects are *cued* to an item in which they must report a potential change. Cues are strictly delayed from the memory array's offset by a minimum of 1000ms, yet this delay may be substantially increased to a 4-second duration. (Phillips 2016; Sligte *et al.* 2008, 2010; Landman *et al.* 2003; Vandenbroucke *et al.* 2011; Block 2011). Importantly, recorded performance suggests that "late" cues are beneficial to subjects' abilities to correctly discriminate changes in any one of 8 items presented up to 4 seconds earlier (Block 2011, Phillips 2016). In fact, this performance illustrates the existence of a longer-lasting sensory memory store ("fragile VSTM") which exceeds the 4-item capacity of working memory: subjects can correctly report changes in 7 of 8 items at 1 second and 5 of 8 at 4 seconds (Block 2011). Thus, with cue delays of up to 1–4 seconds in the paradigm, any conscious experienced formed after the cue's onset appears to require the construction of *visual images* as Phillips (2011a: 211) points out: "…[in] Landman, Spekreijse, and Lamme (2003) and Sligte, Scholte, and Lamme (2008) [i.e. examples of the change detection paradigm]…[visual] imagery does seem implicated."

Let us refer to the item that is originally presented in the memory array, and in which subjects later report the presence or absence of a change, as the "cued item". Moreover, let us consider the cases in which this item is no longer present in the probe array, and thus, subjects must accurately report a change in a newly presented item that replaces the cued item from the memory array. In previous work

(D'Aloisio-Montilla 2017, 2018a), I have argued that the no-overflow account necessarily implicates subjects to generate a *conscious visual image* of the cued item in order to explain their performance in the change detection paradigm. This conscious image, generated during the top-down attentional transfer of iconic memory content into working memory ("WM"), is seemingly what allows subjects to make an accurate change discrimination. Without the construction of a delayed conscious visual image, subjects are unable to bring the cued item to consciousness *at any point in time* during a trial—that is, neither before the cue (as attention has not yet been deployed) or after the cue (as visual imagery is what supports consciousness without appropriate perceptual stimulation). By contrast, overflow theorists allow for subjects to possess a rich consciousness of nearly all items in the grid (including the item that is later cued) prior to the cue, since this is supported by the construction of a conscious "icon" or percept *outside the focus of attention* while the grid is still visible (Block 2011). This conclusion leads to two fatal problems for no-overflow because (1) subjects that are absent of mental imagery ("aphantasics") appear to perform equally well in the change detection task (D'Aloisio-Montilla 2017), and (2) visual imagery abilities, more broadly, do not correlate with performance in the change detection paradigm (D'Aloisio-Montilla 2018a). Together, these imply that subjects do not make use of visual imagery in the change detection paradigm, which, in turn, suggests that subjects must complete the change detection paradigm with *entirely unconscious* representations on the no-overflow account. Given there exists a range of experimental studies that suggest unconscious working memory is too weak to support change detection performance, I have argued that the no-overflow account is left controvertible (D'Aloisio-Montilla 2017, 2018a; Soto *et al*. 2011; Soto and Silvanto 2016; Block 2011; cf. Phillips 2016).

However, a possible *way out* for the no-overflow theorist is to argue that it is a viable conceptual possibility (although, still, empirically controversial) that subjects are capable of providing an accurate change discrimination with entirely "non-conscious, subpersonal" representations—thus rejecting the *necessary* implication for constructing conscious visual images in the task (Phillips 2011b: 407, 2011a; Cohen and Dennett 2011; cf. Block 2011). This paper forms a reply to such a conceptual possibility to make clear that this no-overflow reply would, instead, *necessarily* implicate the construction of an "unconscious visual image" of the cued item, which itself is still shown to be at odds with empirical evidence and conceptual analysis. This naturally leads to a new argument in favour of overflow that can avoid the baggage of the extra premise that subjects are required to construct a conscious experience of the cued item.

Section 2 explores the literature on visual imagery to outline that both unconscious and conscious visual images can be recruited by subjects in memory paradigms (Nanay 2010, 2013, 2016; Phillips 2014; Zeman *et al.* 2010, Zeman *et al.* 2015; Brogaard and Gatzia 2017; Church 2008). Then, Section 3 revisits the change detection paradigm to argue that on the no-overflow account, if necessary consciousness is denied, then subjects must—at the *bare* minimum—generate an unconscious image of previously presented items (including one of the cued item). This differs from previous arguments,[6] since in the past I have chosen to argue that no-overflow necessarily requires subjects to form a *conscious image* of the cued item at some point during trials (CC) (D'Aloisio-Montilla 2017). Thus, in order to account for performance in the change detection paradigm, *for the sake of argument*, I grant that subjects can generate either an *unconscious* or conscious image in providing an accurate change discrimination. Importantly, given this claim allows for conscious imagery to be an *optional* performance strategy in change detection, we arrive at a highly defensible implication: subjects that choose to generate a conscious image of the cued item should perform better than subjects that only generate an *unconscious* image of the item. Multiple defences are given for this claim, but the most compelling builds from the idea that unconscious imagery representations are *necessarily weaker* than conscious imagery representations, since the former is below the threshold for global broadcasting. For instance, as Schzwitzgabel (2011: 51) notes, "people whose imagery is mostly conscious ought to perform somewhat differently on cognitive tasks than people whose imagery is largely unconscious". However,

[6] I mention the possibility of an argument similar to this in Section 4 (Objections and Replies) in D'Aloisio-Montilla 2018a, and make clear that will form the central claim of a further paper. However, this mention does not incorporate the idea of unconscious visual images.

Section 4 cites *recent* empirical evidence from Jacobs *et al.* (2018) and Keogh and Pearson (2011) to demonstrate that subjects who employ conscious imagery *perform as well as* those those that rely on unconscious images. As a result, the no-overflow account is again refuted. Most importantly, this paper's argument applies to all no-overflow views that argue subjects are capable of performing change detection with entirely *unconscious or generically conscious imagistic representations*.[7] Section 5 replies to a range of potential objections and replies, and lastly, Section 6 concludes.

2 Unconscious visual imagery

Visual imagery typically refers to the mental faculty responsible for generating conscious visual experiences in the absence of appropriate external stimuli (Kosslyn 2006, Nanay 2010, Phillips 2014). A more precise and updated notion of visual imagery has been provided by Nanay (2016: 67) as follows: visual images involve "perceptual processing that is not triggered by corresponding sensory stimulation in the relevant sense modality". Thus, in the change detection paradigm, any conscious experience directly caused by perceptual processing—while the memory array is *still visible* to subjects—should not "count" as visual imagery. This means that, as posited by the overflow account, any conscious percept formed while the array is still visible, as well as any ensuing "conscious icon" of the array's items into the interstimulus interval, is not imagistic in nature (Block 2011: 569; D'Aloisio-Montilla 2017, 2018a). From a representational standpoint, such experiences are likely to be supported by iconic or perceptual representations and *not* imagistic representations perhaps in working memory (Block 2011, 2007; D'Aloisio-Montilla 2018a) By contrast, any conscious experience constructed after the memory array's offset is compatible with this definition of visual imagery, such as those which are available to subjects on the no-overflow ac-

[7] If conscious images support better performance than unconscious images due to a necessary difference in representational strength, then any gist-like conscious imagistic representations are clearly *weaker* than specific imagery representations, and so again, performance should improve for subjects that employ specfically conscious images.

count.[8] With regards to various *kinds* of visual images, it has been argued that visual images can differ in at least three key properties, namely whether they are voluntary or involuntary (Pearson *et al.* 2015, 2011), conscious or unconscious (Nanay 2016, Phillips 2014), and whether they are purely visual or invoke *spatial* representations in the visuospatial domain (Shepard and Metzler 1971, Zeman *et al.* 2010, Pearson *et al.* 2015). For the purposes of this paper, we are interested in defining the representational and functional properties of *unconscious* visual images. A good starting point is the recently labelled condition of "Aphantasia", which refers to the *reported* absence of voluntary, conscious visual imagery abilities (Zeman *et al.* 2010, Zeman *et al.* 2015, D'Aloisio-Montilla 2017, Jacobs *et al.* 2018). Aphantasics have been shown to *perform equally* to regular imagers in a range of memory paradigms, thus leading to a potential hypothesis that aphantasics recruit unconscious visual images that do not enter personal-level processing: "there is, in principle, an alternative possibility—that … "blind imagination" depends on the processing of intrinsically visual representations that no longer enter consciousness depends on the processing of intrinsically visual representations that no longer enter consciousness" (Zeman *et al.* 2010: 154, Phillips 2014). Moreover, given the apparent *similarity* of the recorded performance in aphantasics and regular imagers, it seems plausible that aphantasics could similarly recruit imagistic representations that remain unconscious on account of their cognitive position, such as their confinement to sub-personal systems (Phillips 2014; Keogh and Pearson 2014, 2011; Zeman *et al.* 2010). It is important to note, however, that some behavioural differences in task strategies have been recorded in certain imagery tasks (Pearson *et al.* 2015).

There are, therefore, stronger arguments for positing the existence of unconscious visual imagery. First, there is a host of evidence to suggest that visual imagery and visual perception are similarly related faculties (Brockmole *et al.* 2002; Nanay 2010, 2016, 2013). Given it is a widely accepted claim that unconscious perception exists,[9] then there is no clear reason to reject the existence of

[8] This difference between the overflow and no-overflow account is directly addressed in Section 6.

[9] I do not mean to say that unconscious perception is of the same fundamental

unconscious imagery (Nanay 2016, 2010). In fact, if imagery and perception are agreed to be similar in certain faculties, then a host of empirical evidence is unlocked to support the existence of unconscious visual imagery representations (Bartolomeo 2008, 2002; Cichy *et al*. 2012; Ishai *et al*. 2000). For instance, Ishai and colleagues (2000) were able to show the sensory trace of images and precepts are similar, and Cichy and colleagues (2012) evinced images and percepts recruit overlapping visual representations. In other words, the claim that visual imagery is *necessarily* conscious need only be true if it is solely defined in terms of *phenomenology* (visual images are typically *experienced*) and not through its supporting representations or functional roles. Together, the role and representation of imagery might be described as the "format" of visual images allowing us to gain a stronger grasp of their unconscious kind (Phillips 2014: 285). As Phillips (2014: 285) notes "Imagery in [the] representational sense is clearly not conscious by definition." Thus, one conception of visual images allows us to separate its "experiential" aspect from its "representational" content (Phillips 2014: 292, Thomas 2012). This means that unconscious visual images are the same as conscious visual images in terms of the underlying representation but not in their experiential status. As a result, one reason that there seems to be a difference in the performance of imagers and non-imagers in certain paradigms could arise from the conscious status of the visual images: although the representations underlying the visual images can be identical, their phenomenology can vary from subject to subject thus accounting for some personal-level differences in aphantasics.

There are good empirical reasons to believe that unconscious imagistic representations can be recruited by aphantasics and, on occasion, regular imagers. Zeman and colleagues (2010: 154) have argued that aphantasia is caused by "a functional disconnection" between the frontal workspace and posterior regions responsible for perhaps constructing imagery representations. This would allow us to frame unconscious visual images as those that do not enter personal-level processes or a frontal workspace, but nonetheless are supported by isolated imagistic representations (Phillips 2014). Second, such an

---

kind as conscious perception (i.e the "SFK" claim). See D'Aloisio-Montilla 2018b for more.

account of unconscious imagery appears be validated by Schwitzgebel's (2011) account of non-imagers (i.e. aphantasics), since he suggests that "non-imagers" might construct images which they lack introspective awareness of (Pearson *et al.* 2015). This could imply that "unconscious images" might instead be *non-introspectable* conscious images. For the purposes of this paper's argument, if unconscious images are the absence of introspection to conscious images, then this still suffices to refute the no-overflow account, although it seems that the empirical evidence favours the claim that unconscious images are (experientially) *unconscious*. This is because, as Phillips (2014: 292) points out himself, subjects who recruit unconscious images will necessarily perform differently irrespective of whether they are caused by representational or introspective differences: "people whose imagery is mostly conscious… do perform very differently. For considered at the personal-level, the performances of the imager and non-imager are grounded and justified in fundamentally different ways". Lacking the ability to introspect conscious images should lead to worse task performance in imagery-related paradigms, even if this arises from poorer *metacognitive* judgement (Pearson *et al.* 2015). Thus, performance differences between unconscious and conscious imagery would occur in the case that unconscious imagery is caused by (1) lacking introspective access to conscious images or, (2) imagistic representations remaining outside of experience (Phillips 2014).

Moreover, recent findings on aphantasia have evinced clear *behavioural differences* in how regular imagers and non-imagers complete tasks, indicating that the difference between introspected conscious images and non-introspected conscious images *alone* would not suffice to reject the existence of an unconscious experiential type. (Keogh and Pearson 2011, 2014; Jacobs *et al.* 2018). As Nanay (2016: 68) makes clear, "Whatever these experiments say about mental imagery they must say about mental imagery that is not necessarily conscious as these experiments are behavioural experiments and the reasons for inferring the exercise of mental imagery are not introspective ones but they come from the timing of the subjects' responses." For instance, in a series of memory tasks, an aphantasic "MX" performed to the same standard as regular imagers but differed in the speed of response, implying stronger representational and functional differences between imagers and non-imagers, and not just introspective differences.

Given this paper's inquiry is of a philosophical nature, the best argument for unconscious imagery should come from a theoretical standpoint. Nanay (2010) and Church (2008) have provided independent philosophical arguments for the existence of unconscious visual imagery. For instance, Nanay (2010: 239) has argued that "mental imagery is indeed a necessary ingredient of perception itself" thereby allowing for the existence of unconscious visual images on the basis that unconscious perceptual processing is apparent in subjects. The argument stems from *amodal perception*, in which subjects represent parts of an object that are not currently visible to them. Nanay argues that given the objections to alternative accounts of amodal perception, the best explanation of amodal perception is to argue that we perceive these parts with images, thereby arguing that unconscious imagery *necessarily* exists. I will not go into further details regarding this argument given that it is not directly relevant to this paper's inquiry. Instead, the next section provides an empirical account of how unconscious images may be constructed in the change detection paradigm. In sum, Phillips, a no-overflow theorist himself, has acknowledged the possibility of unconscious visual imagery (Phillips 2014), and given recently collected evidence on the existence of aphantasia and non-imagers more broadly, it is clear that unconscious images are *no more exotic* than conscious ones.

As a brief aside, the possibility of "superblindsight"—that is, an ability to have *direct* access to perceptual information without an accompanying phenomenal experience—is an interesting parallel to the condition of aphantasia (Block 1995: 233, Phillips 2014, D'Aloisio-Montilla 2017). Phillips (2014) has pointed out that "nonimagers" (i.e. aphantasics) are comparable to *super-blindseers* in the sense that in certain imagery paradigms, they perform identically to "super imagers" despite lacking any phenomenology or reported conscious experience of visual images (Block 1995). This parallel causes a unique problem for the no-overflow view that I will briefly sketch here, but this should form the basis of a further inquiry. In short, if superblindsight were to exist, then a threat from *epiphenomenalism* appears to be a natural outcome because conscious experience cannot play a key functional role in the reporting of personal-level perceptual judgements (Phillips 2014, Block 1995). However, given this paper will show that *conscious* visual imagery gives no advantage to subjects completing the change detection paradigm, then it appears that the no-overflow view is committed to one of the following claims: either (1) unconscious visual imagery is as equally useful as conscious visual imagery, which, in turn, leads to the undesirable threat from epiphenomenalism as can be drawn from superblindsight, or (2) non-imagers should perform worse in the paradigm as they maintain weaker representations, but this conflicts the empirical evidence as the following sections will show (Keogh and Pearson 2014, Jacobs *et al.* 2018). Although Phillips (2014) rejects the possibility of epiphenomenalism in the case of absent conscious imagery, on the grounds that non-imagers and super imagers may vary in the experiential but not representational status of visual imagery, the *burden of proof* would be on the side of no-overflow to provide an empirical and theoretical justification. This is because this account requires both non-conscious *perceptual* and imagastic processing during change detection.

3 Unconscious imagery and change detection

This section will demonstrate that the construction of unconscious visual images, if not conscious images, can be *necessarily* implicated in the change detection paradigm on the no-overflow account. In the paradigm, it is clear that on the cue's presentation, either the gist-like, generically conscious or unconscious contents of sensory memory are attentionally captured and prepared for change discrimination (Kouider *et al.* 2010, Phillips 2016, cf. Block 2011). Thus, while subjects might have a *vague impression* of the (later) cued item while the memory array is visible, it is only after the post-stimulus cue that conscious access can occur on the standard no-overflow account (Carruthers 2015, Cohen and Dennett 2011, Gross and Flombaum 2017, Deahene *et al.* 2006). If attention were prematurely directed prior to the cue's arrival, then subjects would be unable to accurately capture the cued item's representations unless they had predicted the cue's target *by chance*. This is because drawing attention to sensory memory inhibits the further retention of unattended content (Astle *et al.* 2012, Gressmann and Janczyk 2016, D'Aloisio-Montilla 2017). Thus, focal attention can only be directed following the cue's onset so that it may facilitate sensory representations being transferred into working memory[10] in preparation for conscious access. This means that given that the account assumes attention is necessary for consciousness, subjects can *first* generate a (specifically) conscious experience of the cued item after a delay of 1–4 seconds from the array's offset.[11] Since Kosslyn (2006) has demonstrated that conscious visual images take around 1–1.5 seconds to reach full effect, the upshot is that subjects are only able to generate a conscious or unconscious image of the cued item on cue presentation. Other than visual imagery, there is no agreed upon type of visual representation that can *potentially* support consciousness so long after the perceptual stimulation has ceased. This is because, *ipso facto*, visual images are the modality which support conscious experience not triggered by *currently available* perceptual stimulation (Nanay 2010, 2016; Kosslyn 2006; D'Aloisio-Montilla 2017, 2018a). It is true that subjects might have a fragmentary impression of the memory array's items while it is visible and therefore not *solely* rely on visual imagery to have a generic phenomenology of the grid. However, our argument hinges on how subjects form a *specific* phenomenology of the cued item consistently[12] during trials, and this requires the attentional processes on cueing.

Researchers have implicated the use of visual imagery in the transfer of contents from sensory memory into working memory when this transfer is supported by a "retro-cue" (Dijkstra *et al.* 2017, Griffin and Noble 2003, Pertzov *et al.* 2013, Lee and Baker 2011, Albers *et al.* 2013). As Lee and Baker (2016: 8) note "In comparing results

[10] Working memory raises questions largely beyond the scope of this paper. First, there is much confusion between working memory and visual short-term memory ("VSTM"), yet the latter is generally accepted to be able to maintain representations for a longer timescale. Second, it is unclear whether or not working memory *always* entails conscious access on the no-overflow view given that some theorists allow for unconscious representations to be maintained in working memory.

[11] Notice this is not the case with the overflow position because subjects can experience the cued item with a *percept while the memory array is encoded in (non-attended) sensory memory*. In other words, subjects can form a rich experience while the memory array is visible.

[12] The case in which subjects, by chance, direct attention to the later cued item prior to the offset of the stimulus if not of use here, since it will not account for *consistent* task performance.

from working memory with those from mental imagery it is worth noting that working memory paradigms involving a retro-cue, which requires the retrieval of previously presented information, are not that dissimilar from the paradigms used in mental imagery." There are a number of reasons for this proposed similarity. First, the "additional activity" seen in posterior activations following the onset of a retro-cue bare a similar neural marker to what is seen when subjects generate visual images from recently acquired perceptual content (Griffin and Noble 2003, Farah *et al*. 1988). Second, there is a growing body of empirical evidence to suggest that working memory and visual imagery share *common internal representations* (Tong 2013, Albers *et al*. 2013, Borst *et al*. 2012, Barsalou 2008, Kosslyn and Thompson 2003). For instance, Borst and colleagues (2012) showed that visual imagery and working memory representations are susceptible to similar types of masking and interference. Moreover, working memory and visual imagery appear to activate similar brain regions (Albers *et al*. 2013, Salzar *et al*. 2012). In fact, collected evidence for shared representations between imagery and working memory has recently led to a developed theory that the experience of working memory content must take the form of "mental images" that are supported by distinct "copy" imagery representations (Jacobs and Silvanto 2015: 513). Given that retro-cues are directly responsible for prompting the attentional capture of sensory memory content in the change detection paradigm, these findings suggest that any conscious experience arising on the cue onset is likely to partially recruit common imagery and working memory representations. Thus, a proposed overlap between imagery and working memory supports the claim that subjects can (voluntarily) construct a visual image of the cued item in the change detection paradigm.

The work of Albers and colleagues (2013) sheds further light on the conscious experience that can be constructed on cue presentation in the change detection paradigm. In Albers and colleagues' (2013) study, participants were cued to one of two serially presented gratings following their offset. Importantly, there was a 400ms interval between the second grating and the cue, and so the cue in this paradigm satisfies the definition of a "retro-cue". Next, a task cue determined whether subjects had to maintain the cued grating in working memory or generate a visual image of the cued grating as rotated in a particular angle and direction.[13] After a 10-second retention interval, participants were then shown a probe grating and they were asked to report in which direction the probe had been rotated with respect to the stimulus they had kept in mind (Albers *et al*. 2013). As a result, in trials where subjects were asked to form a visual image of a grating on the onset of the cue, subjects were required to generate a *delayed image* at the same time that they might for the cued item in the change detection paradigm (D'Aloisio-Montilla 2017, 2018a, 2018b). Collected results suggested that activity patterns for working memory and imagery trials were similar, since it was possible to decode the orientation of the stored image of the grating in V1-V3 (Albers *et al*. 2013). This suggested that a "common internal representation" was responsible for working memory and visual imagery during trials, with the further upshot being that "visual cortex acts as a blackboard that is used during both bottom-up stimulus processing and top-down internal generation of mental content" (Albers *et al*. 2013: 1431, Lee and Baker 2013). This adds strength to the claim that subjects typically employ conscious visual imagery of the cued item the change detection paradigm.

First, Albers and colleagues' paradigm demonstrates that subjects can generate and manipulate regular "visual images" when prompted by a retro-cue, and this is exactly what seemingly occurs in the change detection paradigm when the cue is presented (Albers *et al*. 2013). In their task, subjects were required to use the image as a means of performing mental rotation. Second, the no-overflow account explicitly requires some representations of the cued item to be encoded in *working memory* in order for conscious access to occur (Phillips 2016, Carruthers 2015, Cohen and Dennett 2011, cf. Block 2011). Thus, given the observed similarity between visual imagery and working memory representations in Albers and colleageus' (2013) paradigm, and given that subjects must explicitly take advantage of retro-cues in regular change detection, it follows that any conscious experience *first constructed after* the retro-cue's onset in the regular paradigm can also take the form of a voluntary, conscious image on the no-overflow account (D'Aloisio-Montilla 2017, 2018a; Phillips 2011a, 2011b).

Importantly, the claim that subjects may choose to generate a late

[13] 60 or 120 clockwise or anticlockwise.

"visual image" of the cued item in the change detection paradigm is acknowledged by Phillips himself (2011a: 211, 2011b: 403), who is a prominent opponent of overflow. However, there is current disagreement amongst no-overflow interpretations on whether *any* conscious experience (percept or image) of the cued item is *at all* necessary to explain subjects' performance in the task (Phillips 2011a, 2016; cf. Block 2011). As Phillips (2011b: 406) notes explicitly, "note that recognition that there has been a change of orientation is compatible with a lack of conscious recall of the initial memory-array rectangle's orientation". In previous work (D'Aloisio-Montilla 2017, 2018a, 2018b), I have argued that on the no-overflow account, subjects are required to generate a conscious experience of the cued item with visual imagery once attention had been deployed by the cue:

> (CC) Subjects consciously experience a successfully reported, cued item.

I made the case that the no-overflow account must satisfy (CC) on the basis that unconscious working memory is *too weak*—as is evinced by a range of masking paradigms—to account for accurate change discrimination. This is further supported by the existence of the condition of absent visual imagery ("aphantasia") and the unconscious performance limitation in blindsight patients (Soto *et al.* 2011, Zeman *et al.* 2010, Kentridge *et al.* 2004, D'Aloisio-Montilla 2017, Block 2011, cf. Phillips 2016). The idea is that although it is consistent with evidence collected on unconscious memory to suggest that subjects can successfully perform change detection without a conscious experience of *most* of the memory array's items, all accounts should minimally require subjects to have consciousness of the *cued item* at some point in the task. This is because a rejection of (CC) would be highly controversial given the wide array of empirical evidence that has so far been collected on unconscious working memory (D'Aloisio-Montilla 2017, 2018a; Block 2011; Stein *et al.* 2016). For instance, no evidence has suggested that subjects can perform a task as complex as remembering up to 8 items for up to 4 seconds without some conscious experience required (Block 2011). Moreover, the requirement of (CC) does not beg the question in favour of overflow because the opposing view is able to permit subjects forming a *delayed* conscious experience of the cued item on the onset of the retro-cue—once attention has been directed at sensory representations of the item. This section has again suggested that this possible conscious experience must be *imagistic* in nature. However, unlike my previous arguments for overflow, this paper will not require the no-overflow account to necessarily satisfy (CC). Herein, we will therefore assume that, *for the sake of argument*, even though subjects can only form a delayed conscious experience of the cued item with a *visual image*, conscious experience of the cued item is *not necessary* to explain normal performance in the paradigm.

It is useful to contrast the no-overflow appeal to unconscious abilities with the *rich* phenomenology that overflow theorists argue is available to subjects in the paradigm. As stated, the overflow position maintains that subjects possess a *detailed* conscious experience of the memory array's items (Block 2011; D'Aloisio-Montilla 2017, 2018a, 2018b; Bronfman *et al*. 2014). This means that subjects are capable of forming a conscious percept of a still visible memory array, which includes the specific details of *all or almost all* of items, including the details of the item that will be later cued. This conscious experience of the (later) cued item is said to be "long-lasting" and extends into a *post-stimulus* timescale (Block 2007, 2008). And so, in the 1-to-4-second interval between the memory array's offset and retro-cue's onset, the overflow account holds that subjects maintain a phenomenology of almost all of the memory array's items (Block 2008, 2011). This is supported by the iconic and "fragile VSTM" representations held in sensory memory, and is maintained without the focal attention, or working memory capture, that is required for cognitive access. Accordingly, the overflow account maintains that on the onset of the retro-cue, subjects capture cued representations into working memory in preparation for discrimination and report, but ascribe *no phenomenological importance,* whatsoever, to the retro-cue. This is because on the overflow account, subjects are *already* conscious of the cued item at the time of retro-cue's onset. As Block (2011: 571) makes clear, "the overflow argument takes no stand on whether or not the cue erases that conscious icon".

As such, the overflow account does *not at all* implicate the necessary employment of late-stage visual imagery in the change detection paradigm*,* since subjects can simply maintain an already conscious icon on the retro-cue's arrival.[14] This dramatically differs in the case of nearly all[15] possible versions of the no-overflow argument, since we arrive at the following disjunction about the phenomenology available to subjects *after the presentation of the retro-cue* (1) if attention, *by chance*, is directed at the later cued item while the memory array is available, then subjects can possess a specific consciousness of the cued item in some trials but this *cannot* account for consistent task performance (2) on the retro-cue, subjects can generate a delayed, *specifically* conscious visual image of the cued item for the *first time*, or (3) on the retro-cue, subjects can continue to maintain a gist-like consciousness represent the cued item below the threshold for specific phenomenology—that is, without including the item's identity and details. With regards to the latter claim, the cued item might remain in a persisting gist-like or "generically" conscious experience—as higher order theories are likely to posit a shift from gist-like representation to a specifically conscious image in working memory (Brown 2014, Kouider *et al.* 2010, cf. Phillips 2016)—or the item can remain entirely outside consciousness as Phillips (2011a, 2011b) or Cohen and Dennett (2011) suggest. Given this clear difference in *phenomenological timing* proposed by each side of the overflow debate, we should therefore expect to see the following behavioural differences in change detection performances even if phenomenology of the cued item is not a necessary requirement for the task: *subjects who "choose" to construct a delayed, specifically conscious experience of the cued item, taking the form of a conscious visual image, should perform to a higher standard than those that construct or make use of an unconscious visual image of that item on the no-overflow account*. The remainder of this section will justify this relatively uncontroversial claim (as after all, unconscious images are, ipso facto, weaker than conscious images).

It is natural to suggest that constructing a conscious visual image requires stronger representational strength and quality, and so employing conscious imagery should improve the fidelity of stored content and thus task performance (Schwitzgebel 2011; Faw 2009, 1997). As Faw (2009: 20) argues, "non-imagers and even some clinical non-imagers who can still perform (if less well) the tasks

[14] This is revisited in Section 4, where counterarguments are addressed.

[15] See footnote 1 for an acknowledgment of Carruthers' (2015) position.

usually related to imagery…might form some type of 'entry level representation…or form images at a 'subliminal level…incapable of conscious retrieval and reporting"". Taking Faw's subliminal framing and Schwitzgebel's (2011) claims on unconscious images harbouring "different performance", then we can quite clearly think of unconscious images as being below a neural "threshold" such that they do not achieve the *same strength* as conscious images. For instance, conscious images can be globally broadcasted or available to "personal-level" processing in a way that unconscious images cannot be given our current conceptions of working memory (Phillips 2014: 292). As a result, it is very difficult to avoid the consequence that subliminal imagers—that is, those that make use of unconscious images—should perform to a poorer standard in light of their images being, by *empirical necessity*, weaker than conscious images. In fact, Phillips (2014: 292) has independently claimed that , "people whose imagery is mostly conscious… do perform very differently… the performances of the imager and non-imager are grounded and justified in fundamentally different ways", and so there is left a clear ambiguity as to whether "grounded" includes necessary differences in the neural threshold required for unconscious (potentially subpersonal) images to be recruited. Most clearly, Schzwitzgabel (2011: 51) has explicitly mentioned that "people whose imagery is mostly conscious" must "perform…differently" to subjects that employ unconscious images—that is, whose "imagery is largely unconscious".

Thus, the no-overflow position *uniquely* allows for the construction of a delayed conscious image of the cued item to lead to a significantly *improved* performance in the change detection paradigm. This is not the case on the overflow account, for subjects already construct a conscious percept and icon of the cued item prior to the cue. Moreover, the claim that conscious imagery leads to better performance is clearly compatible with the no-overflow interpretations that take a supportive stance on (CC), viz. that subjects must consciously experience a cued item. This is because this section has demonstrated that a conscious visual image is the *only* type of late experience that can be generated (Block 2011; D'Aloisio-Montilla 2017, 2018a). Consequently, if subjects were unable to generate a conscious visual image of the cued item on cue presentation, then their performance abilities would be dramatically attenuated *ex* *hypothesi*. Yet a performance increase is compatible with all versions of the no-overflow argument that posit subjects can complete change detection *without* any conscious experience of the cued item—that is, through entirely relying on unconscious imagery. We have already demonstrated that visual imagery and working memory are likely to recruit a *common set* of internal representations. Thus, we can think of all versions of no-overflow, including those that reject (CC), as being compatible with a different claim CC*:

> (CC*) Subjects must generate either conscious or unconscious (working) memory representations of the cued item on the onset of the retro-cue. Given the apparent overlap between working memory and imagery representations, we might likewise choose to say that subjects must generate either a conscious or unconscious visual image of the cued item.

To be clear, a subject constructing an unconscious "image" of the cued item simply implies that subjects generate or maintain unconscious imagistic *representations* of the item. Thus, (CC*) does not require working memory representations of the cued item to be *conscious*, and so it follows that all no-overflow positions are compatible with (CC*). This is because these interpretations can satisfy (CC*) by maintaining that subjects generate *unconscious* imagastic representations of the cued item when its encoding is transferred from sensory memory to working memory on the cue's presentation. We can therefore choose to frame Phillips (2011a) as positing that subjects rely on an "unconscious visual image"[16] of the cued item in completing the change detection paradigm (Brogaard and Gatzia 2017: 4, Nanay 2010, Church 2008, Nanay 2016, Phillips 2014). Notice that if Phillips were to argue that a delayed unconscious visual image of the cued item cannot be supported by overlapping *working memory* representations, then this does not escape the claim that unconscious performance should be less impressive than conscious performance. This is because we can still attribute subjects as generating an unconscious visual image even if its underlying representations are distinct from working memory. More formally, let us suppose Phillips that rejects the link between unconscious imagery and unconscious

[16] An "unconscious visual image" is simply unconscious imagistic representations.

working memory on account of working memory partially requiring *personal-level* processing, whereas other kinds of unconscious representations only invoke *sub-personal* systems (Phillips 2014). Even so, this does nothing to avoid the claim that subjects can only form a delayed, specifically *conscious* experience of the cued item with a *conscious* visual image, and so subjects that rely on conscious images should perform better than subjects that rely on unconscious images, irrespective of whether unconscious images are encoded in working memory or are indeed sub-personal. There is no reason to assume unconscious mental images are necessarily personal, even if conscious images are. This is because the evidence presented earlier in this section about visual imagery does not *itself* require common conscious and unconscious representations. Thus, it does not matter where unconscious images are *stored*, but rather that unconscious imagery of the cued item is *weaker* than any conscious imagery of that item. I have chosen to propose a link between unconscious working memory and unconscious visual imagery representations because it is gaining traction in the literature, and it unifies the no-overflow account around a common claim on the role for conscious *or* unconscious representations in the change detection paradigm.

We can therefore reframe a no-overflow rejection of a (CC)-like claim as the new claim that subjects must rely on an unconscious visual image of the memory array throughout the entire task in order to perform well: "subjects may recognise change in the probe-array despite never having enjoyed conscious experience of the relevant rectangle in the memory-array" (Phillips 2011a: 406). However, any no-overflow theorist will find it very difficult to argue that completing the change detection paradigm with an unconscious visual image should carry the *same benefit* as performing it with a conscious visual image. This would suggest that a subject generating a conscious image of the cued item carries no added benefit to "guessing" from unconscious (or implicit) working memory representations. This, in turn, implies that unconscious representations in working memory should be as useful to subjects as conscious representations, which clearly contradicts a natural increase in representation quality and precision found in storing items consciously (see Stein 2016 for updated review, Faw 2009, Schwiztgabel 2011). To put it plainly, if subjects choose to generate a conscious image of the cued item, then given that image is compatible with the no-overflow position, there is *no good reason* to then suggest that subjects do not perform to a better standard than when they rely on a purely unconscious strategy. Stein *et al*. (2016: 3) summarize of the recent literature on unconscious vs conscious performance more broadly as follows, "Thus, although recent studies on non-conscious WM opened an exciting new avenue for research on the interplay between consciousness and WM, it would be premature to revise our current understanding of a tight link between WM and conscious awareness.". Moreover, Phillips (2011b: 406) seemingly acknowledges the fact that conscious visual imagery is key to explaining the *most successful performance* in the change detection paradigm in the below excerpt, therefore implying that conscious imagery necessarily leads to better performance than unconscious imagery:

> Block's most persuasive phenomenological appeal is to reports that subjects who are successful in the retro-cue condition construct images to facilitate performance (2008: 309).[30] As Block notes, such constructed [conscious] images are also implicated in temporal integration tasks (Brockmole *et al*. 2002). This may partly explain why training improves performance in the task…Prima facie, then, [conscious] imagery-based performance in the retro-cue condition can also be accounted for in terms of implicit memory effects. Such effects can be quite striking. We should not then be surprised if accurate [conscious] images can be formed by self-prompting subjects in the absence of conscious awareness of the initial memory-array.

The first upshot is that we should not be surprised if subjects can learn to generate conscious images of the cued item *even if* that item had been entirely unconscious while it was encoded in sensory memory when the memory array was still visible. This implies that subjects can draw from implicit memory representations if they are to generate a delayed conscious image, which is exactly what we had assumed the no-overflow position required in this section. Second, since Phillips notes that most "persuasive" overflow argument is to appeal to successful subjects as relying on conscious "images" to facilitate performance, then he is implicitly supportive of the claim that employing conscious imagery (once top-down attention is clearly deployed) improves performance in the change detection paradigm. Specifically, he acknowledges the following as key: "subjects who are successful in the retro-cue condition construct images to facilitate performance". Thus, regardless of whether change detection turns out to be "compatible with a lack of conscious recall", it is difficult for any no-overflow theorist to reject the claim that subjects perform more successfully when they choose to generate conscious images of the cued item (Phillips 2011a). This is because an imagery-based strategy is the only means by which a subject can form *any* (specifically) conscious experience of the cued item throughout the entire task. Moreover, any specific image of the cued item should lead to better task performance than gist-like or generic images of the cued item because of an *empirically necessary difference in the representational strength* of the underlying encoding of gist-like representations. For example, conscious images can be globally broadcasted whereas unconscious images cannot, and broadcasting requires an "ignition" of neural activation (Dehaene *et al.* 2006). Thus, even if conscious imagery is not required to explain how subjects can perform to an average performance standard, subjects who voluntarily construct specifically conscious images should perform better than those subjects who rely on gist-like imagery on the no-overflow account.

In closing this section, I will briefly cite empirical evidence to further defend the claim that constructing conscious images should lead to better task performance than any unconscious imagery strategy in the change detection paradigm. A range of studies have demonstrated that the recruitment of conscious visual images improves performance in visual working memory paradigms (Keogh and Pearson 2011, 2014; Brockmole *et al.* 2002; Richardson 2013; Rodway 2006; Marks 1973; Gur and Hilgard 1975; D'Anguilli 2013). First, a modified paradigm by Brockmole *et al.* (2002) requires subjects to retain the details of an array of dots over the course of a retention interval. Collected results demonstrated that when this interval was reduced to a timescale that prevented conscious images from being generated, subjects' performance dropped to just 40–50% accuracy for recalling the correct square on a trial-by-trial basis. This clearly implies that conscious imagery, or at least the construction of late-stage (specifically) conscious representations, is beneficial to task performance when compared with relying on unconscious images (Block 2011, D'Aloisio-Montilla 2018a). As Block (2011: 572) notes, "subjects report that they are generating an [conscious] image and superimposing it on the partial grid on the screen, and their performance confirms their introspective judgments." Second, Keogh and Pearson (2011, 2014) have directly tested the correlation between conscious imagery and working memory performance. They found that subjects who formed conscious images during the retention interval of two working memory paradigms performed to a higher standard than those who did not. This was determined by correlating task performance to self-reported and objectively measured visual imagery abilities (Keogh and Pearson 2014). Their findings are direct evidence in favour of the claim that a conscious, *imagery-based* performance increase should be seen in the change detection paradigm on the no-overflow account: "the current results suggest that the use of the visual buffer, or [conscious] visual imagery, as a strategy during visual working memory results in better performance".

In sum, this section has demonstrated that the no-overflow account uniquely implicates either conscious or unconscious visual imagery in the change detection paradigm due to a difference in phenomenological timing. This, in turn, implies that constructing (specifically) conscious images of the cued item should lead to improved task performance when compared with a reliance on unconscious representations (or "images") of that item. The next section exploits this implication to provide a new argument for overflow.

4 The argument

The only study that has so far *directly* explored the correlation between the employment of conscious (and voluntary) visual images and subjects' performance in the change detection paradigm is that of Keogh and Pearson (2011). In this study, the conscious imagery strength of subjects was objectively measured through a binocular rivalry paradigm (Keogh and Pearson 2011, 2014). In rivalry, a subject is simultaneously presented with two patterns, one to each eye, and only one of these patterns tends to reach consciousness. It has been shown that if subjects generate a visual image of one of the two patterns prior to the onset of the rivalry display, then this pattern has a far higher chance of becoming dominant (Pearson, Clifford and Tong 2008). Measured as a perceptual bias (%), binocular rivalry therefore equates to the strength of voluntary visual imagery, and this can be used as an objective measurement of imagery abilities (more on this later). In the change detection task, subjects were briefly presented with a stimulus of 7 visual gratings for 500ms. After a cue delay of 400ms, a cue pointing to 1 of 7 earlier presented gratings was shown, and subjects were asked to report the direction in which a probe grating had been rotated. This can therefore be seen as a standard change detection paradigm.

Keogh and Pearson (2011) concluded that *no correlation exists* between subjects' imagery abilities and their performance in the change detection paradigm. As they (2011: 4) write, "individuals with strong imagery tended to perform no better on the iconic memory task than individuals with poor mental imagery." More recently, Jacobs et al (2018: 61) showed that "aphantasics" perform to the same standard as regular imagers in a 4-item change detection paradigm, which bares similarity to the regular paradigm appealed to in the overflow debate *except* it does not necessarily implicate the use of fragile VSTM[17]. Thus, in both cases, it is clear that subjects who rely on unconscious images—or more broadly, an unconscious strategy without conscious images—perform to the same standard in the paradigm when compared with subjects that construct (specifically) conscious images. As Jacobs and colleagues (2018: 62) summarize "We also included the change detection task designed by Wheeler and Treisman (2002) to measure visual working memory performance for feature-bound objects... Mental imagery involves the generation of integrated, featured-bound visual images, but single visual features, like color or shape, can be passively and unconsciously stored in working memory without the need to be integrated into object-like representations."

The last section made clear that given unconscious images are *weaker* than conscious images, then *naturally* some performance advantage should be seen in recruiting conscious images if the overflow account is rejected. This means that the no-overflow account is left highly controvertible, since it minimally requires conscious imagery to lead to better performance than unconscious imagery (of the cued item), and yet empirical evidence stacks up against this

[17] This is because the 4 items presented are within the capacity of working memory.

claim. As a result, I will now demonstrate that the no-overflow account cannot offer a simple counterargument for why constructing a delayed conscious image of the cued item does not lead to any registered performance improvement. Recall that, the overflow account makes clear that imagers and non-imagers construct a *similarly detailed* conscious experience prior to the cue when the stimulus is visible, and this is why no comparative advantage is implicated. As a result, we must prove exactly why the conscious imagery abilities measured by Keogh and Pearson (2011) are, at all, implicated in the performance strategy of subjects in the change detection paradigm. This is because an opponent of overflow may argue that change detection performance does not correlate with *most* types of conscious imagery abilities (or conscious imagery abilities *at all*[18]), since it could be argued that most types of imagery abilities (e.g. imagery from long term memory) offer no explicit benefit to performance. As a result, a critic might argue that the ability for subjects to generate vivid images from long-term memory, or episodic memory, provide little or no added benefit to performance in the change detection paradigm. Of course, this would require an argument for why images generated from different memory stores rely on a distinct set of visual imagery processes, yet there seems nothing theoretically wrong with such a claim, and so it is worth further inquiry to strengthen our argument.

To formalize this potential counterargument, it is useful to think of visual images as having three key dimensions: "source", "attentional mode" and "format". The source of an image is the memory store that encodes its content, the attentional mode is what type of attention is responsible for constructing the image, and the format of an image is the sensory modality that it represents. In the change detection paradigm, the source of any later constructed visual image of the cued item is sensory memory (perhaps specifically, fragile VSTM representations leading up to the cue's presentation). The format of this image is likely to fall under both "spatial" and purely visual modalities, since subjects are typically tasked with also recalling spatial properties of the cued item such as its orientation or shape. (Farah *et al*. 1988). We can therefore think of its format as

[18] This claim is addressed in the next section.

"visuospatial" (Baddeley 2003). The fact that imaging the cued item is likely to require two formats of imagery is an important point, because spatial and visual imagery appear to be dissociable in certain contexts (Sima *et al.* 2013, Aleman *et al.* 2005, Pearson *et al.* 2008). Finally, the attentional mode of the visual image is both endogenous (top-down) and exogenous (bottom-up) attention that is occasioned by the cue (D'Aloisio-Montilla 2017, Janczyk and Berryhill 2014). Cues presented in the change detection paradigm are centrally located and so they clearly elicit some endogenous behavioral effects. In other words, subjects voluntarily "choose" whether or not to pay attention to the cue. However, since the cue is externally presented to the subject, it might also partially cause the *automatic* capturing of cued representations, which in turn might also lead to the automatic (or exogenous) construction of a visual image in some cases.

Thus, our overflow argument only requires that an ability to generate visuospatial images from recently acquired sensory memory content (of the cued item) should naturally correlate with *improved* change detection performance on the no-overflow account. This means that we must demonstrate that Keogh and Pearson's (2011) measurement of imagery abilities isolated the *same type of images* that can be of benefit to successful subjects who "construct images to facilitate performance" in the change detection paradigm (Phillips 2011b: 406). First, Keogh and Pearson (2011) used an *objective* task measurement to evaluate imagery abilities in rivalry. This is crucial, because asking subjects to self-report the perceived vividness of imagery *in general terms*, such as through the use of VVIQ (Marks 1973) or other questionnaires,[19] is more likely to measure the strength of images from an autobiographical or long-term memory source. Second, the rivalry paradigm that was used by Keogh and Pearson (2011) to evaluate imagery abilities requires subjects to generate images with a similar source, attentional mode and format to any image of the cued item in the change detection paradigm. In the rivalry paradigm first used by Pearson and colleagues (2008), subjects are

[19] The VVIQ contains questions along the lines of "think of some relative or friend whom you frequently see (but who is not with you at present) and consider carefully the picture that comes before your mind's eye" therefore requiring the use of long-term memory (Marks 1974).

shown the "rivalry display" that contains two rivalry patterns (one to each eye) for 750ms (Pearson *et al*. 2008). On their offset, they are asked to generate a visual image one of the two rivalry patterns for a 10s blank interval; after which the two rivalry patterns are again shown for 750ms. On trial, subjects are asked to report the dominant pattern before and after the blank interval. Pearson and colleagues (2008) noted that in the paradigm, "imagery of the previously dominant pattern led to somewhat higher levels of perceptual stability…, while imagery of the previously suppressed pattern led to much lower levels of perceptual stability."[20]

As a result, it should be easy to see a similarity in the conscious images and measured abilities which subjects use in the binocular rivalry paradigm and the change detection paradigm. As in change detection, subjects in Pearson and colleagues' (2008) rivalry paradigm must construct an image of one of two patterns that are encoded in sensory memory at the start of each trial. Although only one of these reaches consciousness while they are visible to the subject, this does nothing to discount the claim that representations of both the (later) conscious grating and the (later) unconscious grating are encoded into sensory memory. Any appeal to the phenomenology of either grating should be further discounted from a counterargument, since the no-overflow argument takes the stance that sensory representations of the cued item in change detection can at most be "generically" conscious while the memory array is visible. In other words, no-overflow theorists permit that prior to the imaging of the cued item in the change detection paradigm, its representations can remain entirely unconscious as they might in the rivalry paradigm. Moreover, since nothing overrides sensory memory representations until the second rivalry display in the rivalry paradigm, fragile VSTM representations of the first rivalry display can subsist for the early half of the 10-second imagery interval; and this is exactly what occurs with fragile representations of the cued item prior to the onset of the cue in the change detection paradigm.

This means that subjects in both paradigms construct an image

[20] Thus, imagery biased perception in favor of the imaged pattern, and so the bias obtained with imagery in this paradigm is a good indicator of the strength of a subject's imagery.

with the same source, and this indicates the use of overlapping imagery abilities. As an aside, for the rivalry paradigm used in Keogh and Pearson's (2011) study, subjects were not presented with the first rivalry display prior to them generating an image of one of the two. Nonetheless, this poses *no issues* for the claim that similar imagery abilities are employed by subjects in all three discussed paradigms. Pearson and colleagues' (2008) study made clear that removing the initial onset of the rivalry display (in what they called a "randomized-trial design") had no effects on the strength and frequency of the imagery-based rivalry priming seen in the original set-up. Moreover, increasing the background luminance of the display presented to subjects during the 10-second imagery interval inhibited the priming effects in both the 2008 and 2011 rivalry paradigm, again implying that a common imagery ability was employed by subjects in all three tasks[21] (Pearson *et al.* 2008, Sherwood and Pearson 2010, Keogh and Pearson 2011).

Regarding the format of visual images measured in the rivalry paradigm, a separate experimental design showed that rivalry dominance was most apparent when the *orientation* of the rivalry patterns matched the orientation of the previously imagined pattern (Pearson *et al.* 2008). This "orientation-dependence" clearly indicated that the image which subjects generate in the rivalry paradigm has a *spatial* component, as otherwise determining features would be confined to purely visual properties e.g. colour (Pearson *et al.* 2008). This makes clear that both paradigms require the generation of visuospatial images, which is also seen in the fact that both paradigms use oriented gratings as the items that are shown to subjects. Finally, the attentional mode of the image in the rivalry paradigm is also consistent with a partial mixture of endogenous and exogenous attention. This is because in the rivalry paradigm of Keogh and Pearson's (2011) study, subjects were pre-cued with a letter ("R" or "L") to of one of

[21] To make sure that luminance was only interfering with the generation of visual images, and not working memory processes *in general*, Keogh and Pearson (2011) further investigated the effect of introducing luminance in a number working memory task. No interference in performance was observed. Luminance has also been shown to not effect visual attentional processes, a letter working memory task, or be caused by 'dark adaptation' (Pearson and Tong 2008, Sherwood and Pearson 2010).

the two rivalry patterns prior to their onset. Since this external cue is *symbolic*, it likely requires high-level processing mediated by endogenous attention, and lead to other exogenous effects that are onset by externally displayed cues (Weger *et al.* 2008).

And so, putting together the similarity in the source, format and attentional mode of the constructed images in the two paradigms, it is clear that the binocular rivalry paradigm measures a very similar imagery ability to what successful or higher performing subjects can make use of in the change detection paradigm (Phillips 2011b; Block 2008; D'Aloisio-Montilla 2017, 2018a). Thus, we are on firm ground to justify that there is no positive correlation between conscious visual imagery abilities and change detection performance. Although the Keogh and Pearson (2011) study is the only to explore the relationship between imagery and the change detection paradigm that presents at least 7 items to subjects in the memory array, Ng (2011) likewise demonstrated that no correlation existed between a vividness of imagery and subjects' ability to detect a change in a stimulus that contained 4 items. Finally, Berger and Gaunitz (1977) also demonstrated that "good" imagers were no better at detecting changes in subsequently presented pictures than "poor" imagers: Results indicated that subjects rated as 'good' imagers did not perform differently from those rated as 'poor' imagers". As a result, it is clear that the type of imagery ability measured in Keogh and Pearson's (2011) rivalry paradigm is indeed the same imagery ability that is used by high-performing subjects in change detection on account of the clear overlap in the imagery's source, format and attentional mode.

To take stock on where things stand, the no-overflow position implies that subjects who construct a conscious visual image of the cued item should perform to a higher standard than those who rely on unconscious imagery (or representations), since visual imagery is the *only conscious experience* that is potentially available to subjects in the change detection paradigm. However, it is clear that the results of Keogh and Pearson's (2011) study undermine this necessary implication, with there being no observed correlation between imagery abilities and change detection performance. It follows that the no-overflow argument cannot account the observed performance of subjects in the change detection task. This section has therefore demonstrated that the no-overflow account is put under considerable pressure in light of it implicating a necessary performance difference between subjects who construct delayed, (specifically) conscious images of the cued item and subjects who rely on unconscious images.

5 Objections and replies

This section *briefly* considers three objections against the argument I have presented in favour of overflow. The three most likely lines of reply are: (1) unconscious working memory is powerful enough that conscious imagery should not lead to improved task performance, (2) conscious imagery is not *at all* beneficial to subjects in the change detection paradigm, (3) the overflow account is placed at an equal disadvantage, since it might also imply a necessary correlation between delayed conscious imagery abilities and change detection performance. With regards to the first objection, it seems that unconscious representations are "too weak" to explain the memory of up to 5 items for up to 4–5 seconds (Block 2011: 575; D'Aloisio-Montilla 2017, 2018a, 2018b). Through the use of visual masking, Soto *et al.* (2011) prevented subjects from consciously representing the memory array during the change detection paradigm. This meant that subjects were required to solely recruit unconscious representations in order to make their change discrimination. In the paradigm, only one item was presented to subjects contrasting the 7 or 8 typically presented in the memory array. We would therefore expect performance to be of a high standard in this much simpler task if subjects can use unconscious abilities to perform well in the regular version (Landman *et al.* 2003, Sligte *et al.* 2008). However, subjects were only able to correctly detected changes for unconscious stimuli at the slight above chance rate of 55% (Soto *et al.* 2011). It therefore seems unlikely that unconscious working memory (or equivalently, unconscious visual images) can provide any meaningful use to subjects in the regular version of the change detection paradigm.

In fact, Block (2011) uses the poor performance evinced in the masking paradigm to argue that any explanation of the change detection paradigm must be compatible with a (CC)-like claim on necessary phenomenology, viz. that subjects must consciously experience a successfully reported, cued item (D'Aloisio-Montilla 2017). However, Block's stance on the necessary role for consciousness is considerably more *ambitious* than what is required in this paper's argument. The only assumption that this paper requires in is that, on the no-overflow account, generating a conscious image of the cued item necessarily leads to better task performance than when subjects use unconscious representations—not that consciousness of the item is needed to explain "average" or "regular" performance. Put in other terms, the employment of a conscious strategy must lead to better performance than the use of an unconscious strategy; but not, as Block requires, that an unconscious strategy cannot explain average or normal task performance. This renders this paper's argument in favour of overflow potentially more compelling.

Moreover, recent studies on unconscious working memory have applied pressure to Block's position, since they suggest that the fidelity of unconscious working memory, or equivalently, unconscious imagery, might be higher than first thought (Silvanto and Soto 2012, Soto and Silvanto 2016, Rosenthal *et al.* 2016). But there are further reasons to reject Block's broader interpretation of Soto *et al.*'s (2011) findings. As Phillips (2016) notes, the fact that the memory array is masked in Soto *et al.*'s (2011) study is likely to decrease the quality of perceptual representations. Thus, it can be argued that we should expect subjects' performance to be poorer in this paradigm on the basis of poor representation *quality*, and not lacking unconscious abilities. It has also been shown that such differences in experimental methods can lead to substantial performance changes in working memory paradigms (Persuh *et al.* 2016, cf. Stein *et al.* 2016, Phillips 2016). Thus, it seems that the evidence is mostly *inconclusive* regarding the claim that unconscious abilities can or cannot support an *average* or baseline performance standard in the change detection paradigm. However, all we require is that there are virtually no empirical grounds, *whatsoever*, to suggest that unconscious working memory can bring anywhere near the *same added benefit* as high-fidelity conscious representations in the change detection paradigm. This is the only stance that this paper takes on unconscious abilities.

Let us turn to the second objection, viz. that conscious visual imagery is not *at all* beneficial to performance in the change detection paradigm. Sections 2 and 3 thoroughly outlined why visual imagery is implicated in the change detection paradigm on the no-overflow account. Reasons included the timescale of the paradigm, the common properties of working memory and imagery representations, and the fact that conscious imagery has benefited subjects in a range of other paradigms. Fortunately, there exist two studies which strongly imply that imagery directly benefits "change detection" performance (Saad and Silvanto 2013, Keogh and Pearson 2014). In a modified change detection paradigm, Keogh and Pearson (2014) attenuated subjects' abilities to generate conscious images during a 6-second retention interval between the onsets of the memory array and a probe item.[22] The blocking of imagery abilities was achieved through an increase of background luminance presented to subjects (Pearson *et al.* 2008, Sherwood and Pearson 2010, Keogh and Pearson 2011). It was found that the performance "good imagers" was dramatically affected by whether or not they were able to generate conscious images, since performance was inhibited on the blocking of visual imagery (Keogh and Pearson 2011). As Keogh and Pearson (2014: 1) note, "The disruptive selectivity of background luminance suggests that good imagers, unlike moderate or poor imagers, may use imagery as a mnemonic strategy to perform the visual working memory task." In other words, conscious images were clearly *beneficial* to subjects completing a "change detection" paradigm. A similar finding was reported by Saad and Silvanto (2013). They found that when subjects were engaged in a visual imagery task before being presented with a masked stimulus, they were less likely to detect a change in the orientation of that stimulus once a probe stimulus was shown. In fact, there was a *negative correlation* between the self-reported strength of the visual images formed and subsequent accuracy in change detection (Saad and Silvanto 2013). Together, these studies provide even stronger evidence for the claim that conscious visual images can be useful to *highly successful* subjects in the change detection paradigm. However, the most convincing claim here comes from a theoretical standpoint, *as there is no good reason to suggest that forming a delayed consciousness of a previously presented item cannot be useful to improving the representational quality of the cued item, and thus change discrimination.*

[22] In this paradigm, only 4 items were presented to subjects in the memory array and so no retro-cue was required because the number of items presented was in range of working memory capacity. This paradigm is therefore very similar to the the paradigm central to the overflow debate which tests sensory memory capacities except in the presentation of cues.

Lastly, we should briefly defend why the overflow account remains compatible with the absence of correlation found between imagery abilities and change detection performance. Recall that nearly all versions of the no-overflow account necessarily require that prior to the onset of the retro-cue (i.e. attention), there is *no specific* consciousness of the cued item unless attention had been directed to the *later* cued item, by chance. Although the item might be at most "generically" conscious, there is the potential for a *phenomenological difference* before and after the cue. If subjects choose to generate a conscious image of the cued item on the presentation of the cue, then phenomenology of the item can shift from unconscious to "specific" consciousness, or from gist-like or "generic" consciousness to specific consciousness. Yet that this shift in the precision or existence of specific phenomenology does not occur on the overflow account. Overflow theorists posit that subjects maintain a rich, identity-defining conscious "icon" of the cued item throughout the blank "interstimulus interval" of the paradigm (Block 2011, D'Aloisio-Montilla 2017, Landman *et al.* 2003). Given Section 2 defined (conscious or unconscious) visual imagery as occurring in the absence of perceptual stimulation, I would reject the claim that the overflow account implicates visual imagery *at all*, since the *first* construction of a conscious experience of the cued item occurs w*hile the grid is still visible*, and thus while perceptual activity is available to subjects. Thus, visual imagery is not needed for subjects to perform well and construct consciousness. However, if we choose to ascribe the term "imagery" to the overflow account, then overflow theorists allow for a different kind of "image" (that is, an "icon") of the cued item to be constructed while the memory array is still visible *without attention*, whereas no-overflow theorists implicate the use of imagery that is constructed *with attention*. In other words, whether attention is needed in the visual image's construction differentiates the two accounts. For instance, any image of the item on the overflow account is not attentional in the same way one summoned after the cue would be.

Moreover, such an "image" or "icon" is available to all subjects since it results from iconic memory and fragile VSTM encoding outside of focal attention. This holds even for subjects that have difficult in constructing regular visual images from attention once the stimulus is no longer available (as is the case with the condition of aphantasia). This means that representations of the cued item are already "specifically" conscious on the cue's onset, and so whether subjects generate a further, delayed visual image on the presentation of the cue is *irrelevant* to the prior consciousness *of* the cued item (Block 2008, 2011, 2007). This, in turn, implies that constructing a conscious visual image on the retro-cue is not likely to provide any noticeable benefit—or, at least, significantly less benefit—to subjects' performance when compared with the no-overflow account. By contrast, on no-overflow, constructing a delayed conscious visual image with attention is the *only type* of specifically conscious representation that can be constructed, and this paper has shown that conscious representations are clearly more beneficial than unconscious images. In fact, this delayed image fits the standard definition of mental imagery, since it is constructed multiple seconds after the offset of the original grid (Kosslyn 2006). By contast, if the "conscious icon" implicated by the overflow account is labelled an "image", then it violates this definition. Even if it is argued that generating a regular visual image of the cued item might provide some benefit on the overflow account, say, through helping subjects store durable representations after the probe array is onset, it will be *significantly less* useful given subjects already have a persisting consciousness, and is not needed to satisfy (CC) or (CC*).

6 Conclusion

In conclusion, the burden of proof now rests on opponents of overflow to account for the change detection paradigm, since this paper has put the current account under substantial pressure. In short, given it is both empirically and theoretically justified to suggest that conscious visual images should support better task performance that unconscious visual images, and given that the no-overflow account requires the employment of *at least* unconscious images in the change detection paradigm, then the account is unable to explain why conscious images offer *no performance benefit* over unconscious images. Thus, it seems that visual imagery has a larger role in our understanding of consciousness than we once first thought: imagery picks out a phenomenological difference in the timing of conscious experience in the change detection paradigm, with both sides of the overflow debate differing in their stance. Phillips (2011b, 2011a) has already acknowledged a key role for imagery in the overflow debate, and I have recently made another of claims regarding its pivotal role for the no-overflow account (D'Aloisio-Montilla 2017, 2018a, 2018b). As Phillips (2011a: 220) writes, "visual imagery raises large questions…, it would be interesting to know if those with poor or no (self-reported) visual imagery were equally capable of performing Brockmole's task at 1500 ms delays", and this is naturally extended to questions regarding imagery and performance in the change detection paradigm itself. Imagery can therefore can be used to make testable predictions concerning the performance of subjects in partial report paradigms. Most notably, the condition of aphantasia provides a clear method of testing these claims, since these subjects are known to have an *absence* of voluntary visual imagery. As a result, I look forward to *further* investigations on how subjects with differing imagery abilities perform in a range of partial report paradigms.[23]

Nicholas D'Aloisio-Montilla
Hertford College, University of Oxford
Catte St, Oxford OX1 3BW
nicholas.daloisio-montilla@hertford.ox.ac.uk

References

Albers, Anke; Kok, Peter; Toni, Ivan; Dijkerman, Chris; De Lange, Floris. 2013. Shared representations for working memory and mental imagery in early visual cortex. *Current Biology* 23 (15): 1427–31.

Aleman, André; De Haan, Edward; Kahn, René. 2005. Object versus spatial visual mental imagery in patients with schizophrenia. *Journal of Psychiatry and Neuroscience* 30 (1): 53–6.

Astle, Duncan; Summerfield, Jennifer; Griffin, Ivan; Nobre, Anna. 2012. Orienting attention to locations in mental representations. *Attention, Perception, and Psychophysics* 74: 146–62.

Baars, Bernard; Franklin, Stan; Ramsoy, Thomas. 2013. Global workspace dynamics: cortical binding and propagation enables conscious contents. *Frontiers in Psychology* 4: 1–22.

Baars, Bernard. 1988. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press.

[23] I would like to thank Peter Millican, Patrick Butlin and those close to me for discussion.

Baars, Bernard. 2002. The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences* 6: 47–52.
Baddeley, Alan. 2003. Working memory: looking back and looking forward. *Nature Reviews Neuroscience* 4 (10): 829–39.
Barsalou, Lawrence. 2008. Grounded cognition. *Annual Review of Psychology* 59(6): 17–45.
Bartolomeo, Paulo. 2002. The relationship between visual perception and visual mental imagery: a reappraisal of the neuropsychological evidence. *Cortex* 38 (3): 357–78.
Bartolomeo, Paulo. 2008. The neural correlates of visual mental imagery: an ongoing debate. *Cortex* 44 (2): 107–8.
Berger, Göran; Gaunitz, Samuel. 1979. Self-rated imagery and encoding strategies in visual memory. *British Journal of Psychology* 70 (1): 21–4.
Block, Ned. 2014. Rich conscious attention outside focal attention. *Trends in Cognitive Sciences* 18: 445–7.
Block, Ned. 1995. On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18: 227–87.
Block, Ned. 2007. Consciousness, accessibility and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences* 30: 481–548.
Block, Ned. 2008. Consciousness and cognitive access. *Proceedings of the Aristotelian Society* 108: 289–317.
Block, Ned. 2011. Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences* 15: 567–75.
Borst, Gregoire; Ganis, Georgio; Thompson, William; Kosslyn, Stephen. 2012. Representations in mental imagery and working memory: evidence from different types of visual masks. *Memory and Cognition* 40(2): 204–17.
Brockmole, James; Wang, Ranxiao Frances; Irwin, David. 2002. Temporal integration between visual images and visual percepts. *Journal of Experimental Psychology* 28: 315–34.
Brogaard, Berit; Gatzia, Dimitria. 2017. Unconscious imagination and the mental imagery debate. *Frontiers in Psychology* 8(799): 1–14.
Bronfman, Zohar; Brezis, Noam; Jacobson, Hilla; Usher, Marius. 2014. We see more than we can report: 'cost free' color phenomenality. *Psychological Science* 25(7): 1394–403.
Brown, Richard. 2014. Consciousness doesn't overflow cognition. *Frontiers in Psychology* 15(1399): 1–3.
Carruthers, Peter. 2015. Block's overflow argument. *Pacific Philosophical Quarterly* 97(1): 65–70.
Church, Jennifer. 2008. The hidden image: a defense of unconscious imagining and its importance. *American Imago* 65(3): 379–404.
Cichy, Radoslaw; Heinzle, Jakob; Haynes, John-Dylan. 2012. Imagery and perception share cortical representations of content and location. *Cerebral Cortex* 22(2): 372–80.
Cohen, Michael; Dennett, Daniel. 2011. Consciousness cannot be separated from function. *Trends in Cognitive Sciences* 15: 358–64.

Cohen, Michael; Cavanagh, Patrick; Chun, Marvin; Nakayama, Ken. 2012. The attentional requirements of consciousness. *Trends in Cognitive Sciences* 16: 411–7.

D'Aloisio-Montilla, Nicholas. 2017. Imagery and overflow. We see more than we report. *Philosophical Psychology* 30(5): 545–70.

D'Aloisio-Montilla, Nicholas. 2018a. A brief argument for consciousness without access. *Ratio* 31(2): 119–36.

D'Aloisio-Montilla, Nicholas. 2018b. Phillips on unconscious perception and overflow. *Philosophia* 1–14 (Forthcoming).

Dehaene, Stanislas; Kerszberg, Michel; Changeux, Jean-Pierre. 1998. A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America* 95(24): 14529–34.

Dehaene, Stanislas; Changeux, Jean-Pierre; Naccache, Lionel; Sackur, Jéröme; Sergent, Claire. 2006. Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences* 10(5): 204–11.

Dehaene, Stanislas. 2014. *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. New York: Viking Press.

Dijkstra, Nadine; Bosch, Sander; Van Gerven, Marcel. 2017. Vividness of visual imagery depends on the neural overlap with perception in visual areas. *Journal of Neuroscience* 37(5): 1367–73.

Farah, Martha; Péronnet, Franck; Gonon, Marie. Electrophysiological evidence for a shared representational medium for visual images and visual percepts. *Journal of Experimental Psychology: General* 117(3): 248–57.

Faw, Bill. 1997. Outlining a brain model of mental imaging abilities. *Neuroscience and Biobehavioral Reviews* 21: 283–8.

Faw, Bill. 2009. Conflicting intuitions may be based on differing abilities. *Journal of Consciousness Studies* 16(4): 45–68.

Griffin, Ivan; Nobre, Anna. 2003. Orientating attention to locations in internal representations. *Journal of Cognitive Neuroscience* 15(8): 1176–94.

Gross, Stephen; Flombaum, Jonathan. 2017. Does perceptual consciousness overflow cognitive access? The challenge from probabilistic, hierarchical processes. *Mind and Language* 32: 358–91.

Ishai, Alumit; Ungerleider, Leslie; Haxby, James. 2000. Distributed neural systems for the generation of visual images. *Neuron* 28: 979–90.

Jacobs, Christianne; Silvanto, Juha. 2015. How is working memory content consciously experienced? *Neuroscience and Biobehavioral Reviews* 55: 510–9.

Jacobs, Christianne; Schwarzkopf, Dietrich; Silvanto, Juha. 2018. Visual working memory performance in aphantasia. *Cortex* (105): 61–73.

Janczyk, Markus; Berryhill, Marian. 2014. Orienting attention in visual working memory requires central capacity: decreased retro-cue effects under dual-task conditions. *Attention, Perception, and Psychophysics* 76: 715–24.

Kentridge, Robert; Heywood, Charles; Weiskrantz, Lawrence. 2004. Spatial attention speeds discrimination without awareness in blindsight. *Neuropsychologia* 42: 831–5.

Keogh, Rebecca; Pearson, Joel. 2011. Mental imagery and visual working memory. *PLoS One* 6: e29221.
Keogh, Rebecca; Pearson, Joel. 2014. The sensory strength of voluntary visual imagery predicts visual working memory capacity. *Journal Of Visualisation* 14 (12), 7: 1–13.
Kosslyn, Stephen; Thompson, William. 2003. When is early visual cortex activated during visual mental imagery? *Psychological Bulletin* 129(5): 723–46.
Kosslyn, Stephen. 2006. *The Case for Mental Imagery*. Oxford: Oxford University Press.
Kouider, Sid; De Gardelle, Vincent; Sackur, Jérôme; Dupoux, Emmanuel. 2010. How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Science* 14: 301–7.
Landman, Rogier; Sepkreijse, Henk; Lamme, Victor. 2003. Large capacity storage of integrated objects before change blindness. Vision Research 43: 149–64.
Lau, Hakwan; Rosenthal, David. 2011. Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Science* 15(8): 365–73.
Lee, Sue-Hyun; Baker, Chris. 2016. Multi-voxel decoding and the topography of maintained information during visual working memory. *Frontiers in Systems Neuroscience* 10(2): 1–11.
Marks, David. 1973. Visual imagery differences in the recall of pictures. *British Journal of Psychology* 64: 17–24.
Nagel, Thomas. 1974. What is it like to be a bat? *The Philosophical Review* 83 (4): 435–50.
Nanay, Bence. 2010. Perception and imagination: amodal perception as mental imagery. *Philosophical Studies* 150(2): 239–54.
Nanay, Bence. 2013. *Between Perception and Action*. Oxford: Oxford University Press.
Nanay, Bence. 2016. Hallucination as mental imagery. *Journal of Consciousness Studies* 23 (7–8): 65–81.
Ng, Annalyn. 2011. Quantity vs. quality: individual differences in capacity and resolution of visual working memory. https://deepblue.lib.umich.edu/bitstream/handle/2027.42/85308/annalyn.pdf%3Bsequence=1 (Accessed September 2017).
Pearson, Joel; Clifford, Colin; Tong, Frank. The functional impact of mental imagery on conscious perception. *Current Biology* 18(13): 982–6.
Pearson, Joel; Rademaker, Rosanne; Tong, Frank. 2011. Evaluating the mind's eye: the metacognition of visual imagery. *Psychological Science* 22(12): 1535–42.
Pearson, Joel; Naselaris, Thomas; Holmes, Emily; Kosslyn, Stephen. 2015. Mental imagery: functional mechanisms and clinical applications. *Trends in Cognitive Science* 19(10): 590–602.
Pertzov, Yoni; Bays, Paul M.; Joseph, Sabine; Husain, Masud. 2013. Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental*

Psychology: Human Perception and Performance 39: 1224–31.
Phillips, Ian. 2011a. Attention and iconic memory. In *Attention: Philosophical and Psychological Essays*, edited by Christopher Mole, Declan Smithies and Wayne Wu. Oxford: Oxford University Press.
Phillips, Ian. 2011b. Perception and iconic memory. *Mind and Language* 26: 381–411.
Phillips, Ian. 2014. Lack of imagination: individual differences in mental imagery and the significance of consciousness. In *New Waves in Philosophy of Mind*, edited by Jesper Kallestrup and Mark Sprevak. Basingstoke: Palgrave Macmillan.
Phillips, Ian. 2016. No watershed for overflow: recent work on the richness of consciousness. *Philosophical Psychology* 29(2): 236–49.
Prinz, Jesse. 2012. *The Conscious Brain*. New York: Oxford University Press.
Rosenthal, Clive; Andews, Samantha; Antoniades, Chrystalina; Kennard, Christopher; Soto, David. 2016. Learning and recognition of a nonconscious sequence of events in human primary visual cortex. *Current Biology* 26: 834–41.
Saad, Elyana; Silvanto, Juha. 2013. Contrast and strength of visual memory and imagery differentially affect visual perception. *PLoS ONE 8* (12): e84827.
Salazar, Rodrigo; Dotson, Nicholas; Bressler, Stephen; Gray, Charles. 2012. Content specific fronto-parietal synchronization during visual working memory science. *Science* 338(6110): 1097–100.
Schwitzgebel, Eric. 2011. *Perplexities of Consciousness*. Cambridge, MA: MIT Press.
Shepard, Roger; Metzler, Jacqueline. 1971. Mental rotation of three-dimensional objects. *Science* 171 (3972): 701–3.
Sherwood, Rachel; Pearson, Joel. 2010. Closing the mind's eye: incoming luminance signals disrupt visual imagery. *PLoS ONE* 5(12): e15217.
Sima, Jan; Schultheis, Holber; Barkowsky, Thomas. 2013. Differences between spatial and visual mental representations. *Frontiers in Psychology* 4(240): 1–15.
Sligte, Ilja; Scholte, Steven; Lamme, Victor. 2008. Are there multiple visual short term memory stores? *PLoS ONE* 3(2): 1–9.
Sligte, Ilja; Vandenbroucke, Annelinde; Scholte, Steven. 2010. Detailed sensory memory, sloppy working memory. *Frontiers in Psychology* 1(175): 1–10.
Soto, David; Mantyla, Teemu; Silvanto, Juha. 2011. Working memory without consciousness. *Current Biology* 21(22): R912-R913.
Soto, David; Silvanto, Juha. 2016. Is conscious awareness needed for all working memory processes? *Neuroscience of Consciousness* 1: 1–3.
Sperling, George. 1960. The information available in brief visual presentations. *Psychological Monographs: General and Applied* 74(11): 1–29.
Stazicker, James. 2011. Attention, visual consciousness and indeterminacy. *Mind and Language* 26(2): 156–84.
Stein, Timo; Kaiser Daniel; Hesselmann, Guido. 2016. Can working memory be non-conscious? *Neuroscience of Consciousness* 1: 1–3.
Thomas, Nigel. 2012. Mental imagery. *The Stanford Encyclopedia of Philosophy*.

http://plato.stanford.edu/archives/win2012/entries/mental-imagery/ (accessed September, 2017)

Tong, Frank. 2013. Imagery and visual working memory: one and the same? *Trends in Cognitive Sciences* 17(10): 489–90.

Vandenbroucke, Annelinde; Sligte, Ilja; Lamme, Victor. 2011. Manipulations of attention dissociate fragile visual short term memory from visual working memory. *Neuropsychologia* 49(6): 1559–68.

Weger, Ulrich; Abrams, Richard; Law, Mark; Pratt, J. Attending to objects: endogenous cues can produce inhibition of return. *Visual Cognition* 16(5): 659–74.

Zeman, Adam; Dewar, Michaela; Della Sala, Sergio. 2015. Lives without imagery. Congenital aphantasia. *Cortex* 73: 378–80.

Zeman, Adam; Della Sala, Sergio; Torrens, Lorna; Gountouna, Viktoria-Eleni; McGonigle, David; Logie, Robert. 2010. Loss of imagery phenomenology with intact visuospatial task performance: a case of 'blind imagination'. *Neuropsychologia* 48(1): 145–55.