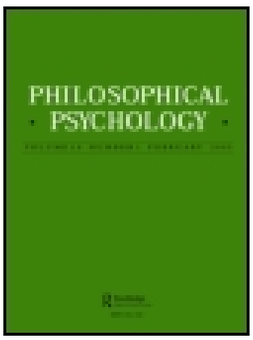

Routledge
Taylor & Francis Group

Philosophical Psychology

ISSN: 0951-5089 (Print) 1465-394X (Online) Journal homepage: http://www.tandfonline.com/loi/cphp20

No watershed for overflow: Recent work on the richness of consciousness

Ian Phillips

To cite this article: Ian Phillips (2015): No watershed for overflow: Recent work on the richness of consciousness, Philosophical Psychology, DOI: 10.1080/09515089.2015.1079604

To link to this article: http://dx.doi.org/10.1080/09515089.2015.1079604

PHILOSOPHICAL PSYCHOLOGY, 2015
http://dx.doi.org/10.1080/09515089.2015.1079604

Routledge
Taylor & Francis Group

No watershed for overflow: Recent work on the richness of consciousness

Ian Phillips

St. Anne's College, Oxford University, Oxford, UK

ABSTRACT

A familiar and enduring controversy surrounds the question of whether our phenomenal experience "overflows" availability to cognition: do we consciously see more than we can remember and report? Both sides to this debate have long sought to move beyond naïve appeals to introspection by providing empirical evidence for or against overflow. Recently, two notable studies—Bronfman, Brezis, Jacobson, and Usher (2014) and Vandenbroucke, Sligte, Fahrenfort, Ambroziak, and Lamme (2012)—have purported to provide compelling evidence in favor of overflow. Here I explain why the data from both studies are wholly consistent with a "no overflow" interpretation. Importantly, when framed purely in representational or informational terms, this "no overflow" interpretation agrees with the interpretations respectively offered by both Bronfman et al. (2014) and Vandenbroucke et al. (2012). The difference only emerges when additional assumptions are made concerning which representations correspond to elements in conscious experience. The assumptions made by overflow theorists are contentious and poorly motivated. However, challenging them simply reopens the original controversy. The upshot is a sobering moral: we still do not know how to move beyond appeals to naïve introspection in establishing the nature and limits of our ordinary experience.

KEYWORDS

Consciousness; Generic Experience; Perceptual Representation; Phenomenal Overflow; Unconscious Perception

1. Background and Overview

A familiar and enduring controversy surrounds the question whether our phenomenal experience "overflows" availability to cognition. According to "overflow" theorists, our conscious experience is rich and detailed, outstripping our cognitive capacities. According to "no overflow" theorists, whereas the world is rich and detailed, our experience of it is limited by what, in principle, we can remember and report. Experimental attempts to settle this controversy have a venerable history, as both sides have sought to move beyond naïve appeals to introspection and offer empirical evidence for or against overflow (Phillips, 2011a). According to many overflow theorists, a crucial advance was made by Sperling (1960). Sperling showed that, although subjects can only report just over four letters from a twelve letter array when asked to report *all* the letters; subjects trained to provide *partial* reports prompted by a post-cue presented within a few hundred milliseconds of display offset can consistently report around three letters from the cued row. Sperling interpreted this "partial report advantage" (the capacity of subjects to report more letters with respect to an immediately post-cued row than with respect to a randomly chosen row in an uncued trial) as evidence that subjects typically enjoy a rich

CONTACT Ian Phillips ian.phillips@st-annes.ox.ac.uk

conscious experience of the array, corresponding to a rapidly decaying "iconic memory." Sperling's idea was that the cue enabled selective transfer from this rich iconic representation into a more limited capacity but durable memory store supporting subsequent report. Adopting an interpretation along these lines, many psychologists and philosophers fête Sperling's work as powerful evidence in favor of phenomenal overflow (e.g., Baars, 1988; Block, 1995, 2007; Burge, 2007; Dretske, 2006; Tye, 2006).

Numerous critics have, however, contested whether this "overflow" interpretation is mandatory. Instead, they contend that Sperling's data are consistent with a view on which subjects only consciously experience as specific letters those items which they can later report as such (Cohen & Dennett, 2011; Phillips, 2011a, 2011b; Stazicker, 2011; compare de Gardelle, Sackur, & Kouider, 2009; Kouider, de Gardelle, Sackur, & Dupoux, 2010). According to these critics, whilst the partial report advantage does reveal that the identities of most or all of the twelve individual letters are *represented*, it does not follow that such representations correspond to elements in conscious experience. In particular, we do not have to think of the post-cue as determining which items from rich conscious experience are stored in explicit memory—we can instead think of the post-cue as "postdictively" determining which unconsciously represented letter identities achieve perceptual consciousness (Phillips, 2011a, 2011b). As well as positing rich *unconscious* representations, "no overflow" accounts must also say something about our experience of uncued rows. One natural proposal is that unreported letters in uncued rows are represented in a *generic*, *gist-like*, or *merely determinable* manner, e.g., as a series of "letter-like forms" as opposed to specific letters (Cohen & Dennett, 2011; Phillips, 2011b; Stazicker, 2011). After all, subjects do believe and report that they saw letters or letter-like forms filling the whole array—they just can't specify the identities of more than three or four.[1]

A notable feature of the disagreement here is that both sides concur when their accounts are framed purely in representational or informational terms. Thus, both "overflow" and "no overflow" theorists recognize rich "iconic" representations of the Sperling-array. What they disagree about is whether these representations correspond to conscious experience.[2] Likewise, both "overflow" and "no overflow" theorists acknowledge the existence of generic or purely determinable *explicit memory* representations of uncued rows. What they disagree about is whether such representations correspond solely to our explicit memory (our experience itself being rich and determinate) or instead reflect the contents of our conscious perceptual experience. The disagreement thus raises the notoriously difficult question of how to bridge between facts about representations in an information processing account of perceptual cognition and facts about conscious experience.

Despite the apparent intractability of these debates, a number of recent experimental papers have claimed to provide compelling evidence in favor of overflow. Two salient examples are Bronfman et al. (2014)—hailed by Block (2014, p. 445) as delivering a "dramatic advance" in favor of overflow—and Vandenbroucke et al. (2012)—cited as evidence by eight of the most prominent researchers in consciousness studies as "evidence that there can be conscious experience without attention or report" (Block et al., 2014, p. 556). Here I explain why both studies are quite consistent with a "no overflow" interpretation. In particular, I explain how, as with earlier responses to Sperling, "no overflow" interpretations can be provided which fully endorse the interpretations offered by both Bronfman et al. (2014) and Vandenbroucke et al. (2012) *when they are framed purely in representational or informational terms*. What this reveals is that both studies only appear to provide evidence in favor of overflow because of certain background assumptions concerning which representations correspond to elements in conscious experience. One such assumption, explicitly defended in Block (2011), is that conscious perceptual representations cannot be generic. A second set of assumptions, explicit in Block (2011) and Vandenbroucke et al. (2012), is that unconscious representations are fundamentally limited in their contents and/or strength. As I argue, these assumptions are poorly motivated and easily challenged by the "no overflow" theorist. Deprived of them, the proponent of overflow inevitably retreats to a direct appeal to naïve introspection. The sobering upshot is that as things stand we do not know how to move beyond such appeals.

My discussion is structured as follows. Section 2 provides details of Bronfman et al.'s study and their argument for overflow. Section 3 offers an alternative "no overflow" interpretation of their core

data, explains why additional empirical considerations adduced do not tell against this interpretation, and concludes that the paradigm only favors overflow on the assumption that conscious perceptual representations cannot be appropriately generic. Section 4 considers this assumption head-on, arguing that Block's explicit defense is unconvincing. Section 5 then turns to Vandenbroucke et al.'s study; sets out how their interpretation of their data as evidence of overflow again turns crucially on a background assumption, this time that so-called "perceptual" representations are necessarily conscious; and again explains why this assumption is controvertible. Section 6 turns to Block's related claim that unconscious representations are too weak or short-lived to play the role required of them by a "no overflow" view and argues that it too is unpersuasive. Section 7 concludes.

2. Bronfman et al. (2014)

A recent study by Bronfman et al. (2014) has been hailed by Block (2014, p. 445) as delivering a "dramatic advance" in favor of phenomenal overflow. In this section, I describe the core data on which Bronfman et al. base their argument for overflow. In the next section, I offer an alternative "no overflow" interpretation of these data, explaining why Bronfman et al.'s additional considerations in favor of an "overflow" interpretation are unconvincing.

Bronfman et al. used a modified Sperling-style paradigm in which subjects were shown a 4 × 6 colored letter-array, with one row pre-cued. The display was followed by a blank interval after which a post-cue indicated to subjects which letter they were to recall and report from the pre-cued row. In some trial blocks, subjects were additionally asked to estimate the color diversity of either the cued or uncued rows. Letters in uncued rows varied between high and low color diversity independently of the color diversity of the cued row. Figure 1 shows a trial sequence. Bronfman et al.'s headline finding—interesting in its own right, and undisputed here—is that subjects could estimate color diversity in uncued rows without cost to letter-recall. This evinces that color diversity can be registered

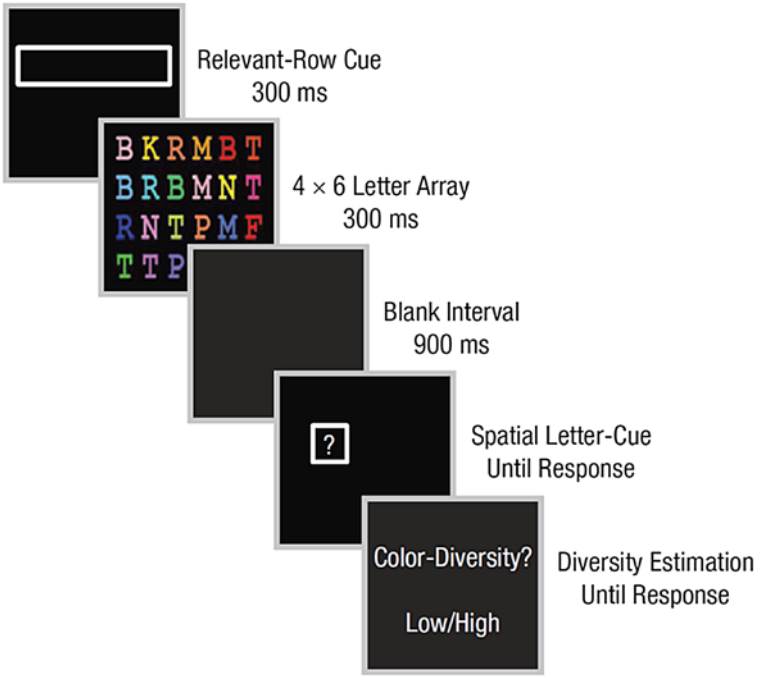

Figure 1. Trial sequence for Bronfman et al.'s experiments 1–4. The final diversity estimation question either refers to cued or non-cued rows depending on the trial block. From Bronfman et al. (2014, p. 1397). Copyright © 2014, Association for Psychological Science. Reprinted by permission of SAGE Publications.

without "focal" attention (i.e., with attention maximally occupied with the cued row) and "cost free" (i.e., without working memory drain).

Bronfman et al. explain their striking finding that letter-memory capacity is unaffected by storing color diversity information by proposing that such information is stored as a one-bit "summary statistic, which rapidly compresses the high-complexity information in the visual display into a binary low/high variable that may be registered and stored while the information itself decays" (2014, p. 1395). However, based on modelling under noise, they argue that this statistic or "ensemble representation" (in contrast to average color) cannot be determined unless the colors of letters are individually and determinately represented. Furthermore, they argue that these specific color representations are conscious. If so, conscious perception is indeed richer than working memory, since its contents include the colors of many or all individual letters in the non-cued rows. These individual color representations are not recruited into the "global workspace" (Baars, 1988) and are not cognitively available for subsequent recall and report (holding fixed the subject's distribution of attention). Thus, if Bronfman et al.'s interpretation of their data is correct, it is understandable why Block should laud their study as marking a "dramatic advance" in favor of overflow.

3. A "No Overflow" Interpretation of Bronfman et al.'s Data

Bronfman et al.'s interpretation can very largely be endorsed by opponents of overflow. In particular, when presented solely as a story about the representational processes involved in color perception and discrimination, they need find nothing to disagree with. It is only in connecting this account with our phenomenal experience that opponents of overflow will call foul. In particular, the opponent of overflow can grant that rich information about individual colors is perceptually registered and then rapidly transformed into a summary statistic. What they will deny is that we must think of the *input* as opposed to the *output* of this process as correspondent to conscious experience. The alternative "no overflow" hypothesis is then that, outside focal attention, subjects have only generic experience based on the output summary statistic (e.g. "diversely colored letter-like forms"), not conscious experience of the colors of specific individual letters (e.g. "cyan-B, orange-K, mauve-T"). Block argues, for instance, that

> ensemble perception of the kind reported by Bronfman reveals that there must have been conscious awareness of specific colors beyond the limits of the global workspace because a trace of that conscious awareness in the form of a diversity judgment can enter the global workspace for free. (2014, p. 446)

According to the alternative hypothesis, the "no overflow" theorist can retort that ensemble perception of the kind reported by Bronfman reveals only that there must have been *representations* of specific colors outside of the global workspace because a summary transform of those representations can enter the global workspace for free. The data do not tell us whether these representations are conscious or not.

Of course, different theorists will find these different interpretations more or less compelling. But, from a neutral standpoint, both look to be eminently reasonable explanations of the data. Moreover, as noted above, the idea of generic, or merely determinable, conscious content is a standard supposition of the "no overflow" theorist (e.g., Cohen & Dennett, 2011; Phillips, 2011b; Stazicker, 2011). As such it represents no new cost to the "no overflow" position.[3] Likewise, the claim that rich perceptual representations are found outside of consciousness is also a standard part of a "no overflow" account (recall the explanation of the "partial report advantage" in terms of unconscious letter representations discussed in connection with Sperling's original paradigm). Thus, again this aspect of the proposal represents no new cost for a "no overflow" theorist. As described in subsequent sections, "overflow" theorists—in particular Block (2011)—have been highly critical of both these aspects of "no overflow" accounts. However, if they are right about this, the "no overflow" theorist is in trouble quite independently of Bronfman et al.'s findings. On the other hand, if, as I argue below, such criticisms are unpersuasive, Bronfman et al.'s basic data pose no new threat to the "no overflow" theorist.

Bronfman et al. themselves offer two further pieces of evidence for the view that individual color representations must be conscious, and so for rich conscious perception. I now explain why neither is

compelling. First, subjects were told to press an escape key if they did not see colors in the uncued rows. Subjects were 93% accurate in detecting catch trials with uncolored rows, and made no false alarms in non-catch trials. Bronfman et al. claim that this makes any explanation which denies the conscious perception of the colors of individual letters "unlikely" (2014, p. 1400). However, the performance of their subjects is entirely consistent with merely generic conscious experience of color outside focal attention. The opponent of overflow need not suppose that there is *no* experience of color without focal attention, only that such experience is generic, not specific.[4] As a result, it is to be expected that when experiencing high or low diversity colored letter-like forms, subjects will not press the escape key; but when experiencing an *absence* of color (an entirely different experience), they will.

Second, Bronfman et al. (experiment 5) show that above-chance color diversity judgments correlate with color visibility ratings using brief, masked presentations. From this they conclude that color diversity judgments cannot be "supported by unconscious color processing" (2014, p. 1400). It is natural to worry that masking conditions do not replicate perception outside attention. But we can bracket this concern. The more fundamental problem is that a correlation between visibility and color diversity judgments is entirely consistent with the "no overflow" account sketched above. On that account, color diversity judgments can be thought of as grounded in generic conscious experiences of color diversity, which in turn are supported by unconscious color processing of specific colors. As a result, the account predicts that where masking degrades the unconscious registration and/or processing of specific colors, subjects will fail to enjoy accurate generic conscious experiences of color diversity based on these unconscious representations and, in turn, will be unable to make accurate color diversity judgments. Making the natural assumption that visibility ratings reflect the quality of conscious generic color diversity representations, it follows that above-chance color diversity judgments will correlate with visibility ratings just as Bronfman et al. find. To treat such a finding as evidence against a "no overflow" account, Bronfman et al. must either ignore or exclude the possibility of generic experiences of color diversity.

Bronfman et al. attempt to strengthen their case that the specific color representations must have been conscious by showing that under slightly stronger masking conditions *dominant* color judgments *can* be made "subliminally" (experiment 6). The difficultly here is that whilst they succeed in showing that reasonably intensive masking does not abolish *all* color processing, they do not show what needs showing, namely that reasonably intensive masking does not abolish color processing *sufficient to produce reliable color diversity representations*, and so to ground accurate diversity judgments. Bronfman et al. are clear that average, and so presumably dominant, color "requires much less perceptual differentiation" than color diversity (2014, p. 1395). Likewise, they provide modelling evidence that average color is far more robust to noise than color diversity. Assuming that intensity of masking corresponds to amount of noise, as masking intensity increases we should expect the abolition of diversity representations well before the abolition of average or dominant color representations. Both sides in the debate will thus anticipate that average or dominant color judgments will be differentially spared under intensive masking.

The only residual issue to explain is why the accuracy of dominant color judgments, unlike that of color diversity judgments, is above-chance even when subjects do not report having seen the colors. Bronfman et al. take this to show that "unlike the discrimination of average color, which can be carried out subliminally, the evaluation of color diversity appears to require some degree of subjective consciousness" (2014, p. 1400). However, even if this is right, this conclusion is still neutral regarding the "overflow"/"no overflow" dispute. After all, there is no reason that "no overflow" theorists should not grant that *generic* diversity representations are associated with consciousness. Their only resistance is to the claim that the *specific* color representations supporting this diversity representation must be conscious. Nothing in Bronfman et al.'s data establishes that claim.

As an aside, it should also be questioned whether Bronfman et al.'s data truly license the conclusion that color diversity representations are distinctive in requiring some degree of subjective consciousness. One ground for doubt here is that Bronfman et al. rely on subjective reports to ascertain phenomenal experience. Yet such reports are notoriously prone to bias and variation across subjects and conditions

(e.g., Macmillan & Creelman, 2005, chapter 2; Reingold & Merikle, 1990). Given that in both average/dominant and color diversity judgments, as subjective visibility increases, so too does accuracy (in addition to their main text, see figure S7 in the supplementary materials), variation in response criteria between the two participant groups could very plausibly suffice to explain the apparent existence of "subliminal" perception in the case of average color and its apparent absence in the case of color diversity. As a result, further investigation is needed before clear conclusions can be drawn from the data under consideration.[5]

In sum: neither Bronfman et al.'s basic paradigm, nor the further evidence which they adduce, provides a compelling case for preferring an "overflow" interpretation of their data. That interpretation seems mandatory only if a "no overflow" interpretation which posits rich unconscious perceptual representations combined with generic conscious experience outside of focal attention is excluded. Below I address in principle objections to such an interpretation which may explain this exclusion. I begin with Block's critique of generic conscious experience.

4. Block's Critique of Generic Conscious Perceptual Representation

"No overflow" interpretations of both Bronfman et al.'s study and Sperling's original partial report task propose that subjects enjoy conscious experiences with generic contents, for instance representing uncued rows as containing "diversely/uniformly colored letter-like forms." Block argues that such generic experiential contents verge on incoherence and lack empirical plausibility. Thus, focusing on the hypothesis of generic letter representations in relation to uncued rows in a Sperling array, Block writes: "It is difficult to understand what it would mean for subjects' consciousness … to consist of conscious representations of a grid of instances of letterness without any specific shape representations" (2011, p. 570). Very plausibly, Block would make the exact same complaint concerning generic color representations, namely, that it is difficult to understand what it would mean for subjects' consciousness to consist of representations of a set of items as (diversely or uniformly) colored without any specific colors being represented.

The grounds for Block's incoherence claim are unclear, however. There is of course no reason in general to doubt the possibility of generic representation. Thus, Block's thought must be that conscious perceptual experience (or some determinable thereof) cannot have generic contents. This explains why Block is at pains to provide experimental support for the idea that we have an "image-like representation" in play in the studies which he is concerned with (2011, p. 573). Here, however, we ought to guard against a fallacy which Block himself drew attention to in an important contribution to the mental imagery debate from the early eighties. The fallacy, which he labelled the "photographic fallacy," is that of thinking that pictorial representation "must be determinate with respect to every visual feature … determinate in every visual respect (i.e., determinate under every visual description)" (Block, 1983, p. 653). The fallacy, Block claimed, was committed by opponents of pictorialism—the view that the representational format of mental images is pictorial. It is a fallacy, he contended, because "for any of a wide class of visual features [including color, size, and shape, see pp. 655–656], one can find (or invent) a type of pictorial representation in which that visual feature need not be determinate" (1983, p. 653). Block's basic point is simply this: pictorial or image-like representation (including photographic representation) does not entail determinate representation; hence, opponents of pictorialism are wrong to argue that mental imagery conceived on pictorialist lines must be determinate in all respects.

Returning to the debate about overflow, Block appears to commit the very fallacy which he once took pains to expose: the fallacious move from image-like to determinate representation. After all, if photographs and other forms of image-like representation can depict merely determinable features across the board (again see Block, 1983, pp. 655–656), what rationale is there for denying that we can make sense of such representation in conscious perception?[6]

Block also argues that generic contents are empirically ill-motivated. In particular, he criticizes the suggestion made by proponents of generic contents that peripheral vision provides an everyday example of generic representational content.[7] Block responds thus: "even if it is true that we can

see movement without shape in the periphery, this would probably stem from specialized motion-detection circuits in cortical area MT/V5 and does not show that we can consciously see generic letterness" (2011, pp. 570–571). This response is puzzling since the suggestion that peripheral vision provides an example of generic representation is not in any way confined to (or peculiarly motivated by) motion perception. Thus, Cohen and Dennett write: "As stimuli move further into the periphery, they gradually lose their color and fidelity … [such that only] certain basic features and statistics are preserved" (2011, p. 360). The concern here is not specifically with motion. Indeed, Cohen and Dennett cite classic work by McConkie and Rayner on peripheral awareness of words in reading—a case obviously germane to Sperling-type tasks. A standard interpretation of McConkie and Rayner's work is that subjects are merely aware of generic letter-like textures outside of focal attention. This explains why subjects fail to notice the replacement of words by strings of Xs when they lie outside fixation (McConkie & Rayner, 1975; Rayner, 1975).

More generally, the "no overflow" theorist can point to a wealth of work on the perceptual representation of "ensemble visual features", i.e. representations corresponding to "any summary statistic that collapses across individual image details" (Alvarez & Oliva, 2008, p. 397). Evidence for such ensemble representations has been found for size (Ariely, 2001; Chong & Treisman, 2003), orientation (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), contrast (Chubb, Nam, Bindman, & Sperling, 2007), location (Alvarez & Oliva, 2008), facial expression (Haberman & Whitney, 2010), as well as, of course, motion (Williams & Sekuler, 1984).[8] As with related work on change and inattentional blindness (e.g., Mack & Rock, 1998; Rensink, O'Regan, & Clark, 1997), the interpretation of this work is controversial. For example, it might be argued that the ensemble representations evidenced by the relevant paradigms do not directly reflect the contents of conscious perceptual experience but only of what is subsequently stored in explicit memory. Again, we see the difficulty of moving between claims about perceptual representations and claims about the contents of consciousness. Nonetheless, as Bronfman et al. themselves accept, such work is clearly consistent with the postulation of "impoverished or blurry experience" (2014, p. 1395). That concession is all that is required to resist Block's case against generic contents.

Bronfman et al. do not themselves explicitly rely on a rejection of generic contents. Their case is founded on the idea that the representation of color diversity "necessitates differentiation between the individual elements' colors" and that this shows there must be more in consciousness than mere ensemble representations of color diversity (2014, p. 1395). However, this takes us to a second set of issues, to be taken up in the remaining sections, namely a background animus against appealing to rich unconscious perceptual representations. For, as I argued in section 3, Bronfman et al. do not show that the representations of individual colors are conscious—and if they are not conscious, they do not evidence overflow.

5. The Limits of Unconscious Representation: Vandenbroucke et al. (2012)

A very explicit appeal to a connection between perceptual contents of a certain kind and consciousness is made in a recent study by Vandenbroucke et al. (2012; see also Vandenbroucke, Fahrenfort, Sligte, & Lamme, 2014). In this section I explain the basic structure of their study and the argument made for overflow on its basis. I then explain why the crucial assumption which underpins it is eminently disputable.

Vandenbroucke et al.'s paradigm is based on a different descendent of Sperling's paradigm developed and popularized by a group at the University of Amsterdam (Landman, Spekreijse, & Lamme, 2003; Sligte, Scholte, & Lamme, 2008; Sligte, Vandenbroucke, Scholte, & Lamme, 2010). The Amsterdam paradigm involves a change detection task and exploits substantially longer post-cue delays than Sperling's iconic cues. In Vandenbroucke et al.'s version of the paradigm, subjects were presented first with a 500 ms *memory display* containing eight figures. A *test display* was then presented in which one of the eight figures had changed configuration on 50% of trials. Subjects were asked to indicate whether there had been a change or not. In both conditions of interest a visual cue indicated the figure

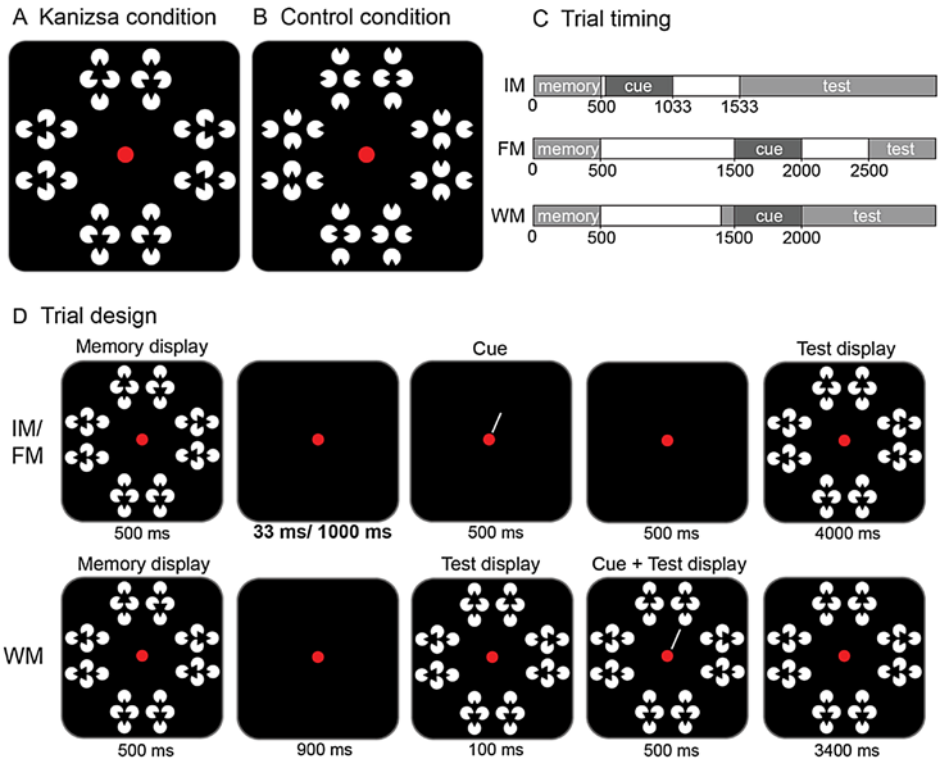

Figure 2. Display configurations and trial sequences used in Vandenbroucke et al. (2012). The reader can ignore the iconic memory (IM) condition. In the example sequences shown, the cued figure in the display changes configuration. Only cued items ever change, and change occurs on 50% of trials. From Vandenbroucke et al. (2012, p. 3). Copyright © 2012, Vandenbroucke et al. Reprinted under the terms of the Creative Commons Attribution License.

potentially subject to change. The cue lasted 500 ms and was displayed at a delay of 1000 ms from memory display offset. In the *working memory* (WM) condition, the test display was presented 100 ms *before* the post-cue; in the *fragile memory* (FM) condition, the test display was presented 500 ms *after* post-cue offset (see figure 2, C and D). In line with other results from the Amsterdam group, Vandenbroucke et al. found that performance was significantly better in the change detection task in the *fragile memory* as opposed to the *working memory* condition. This basic finding has been interpreted as revealing a fragile form of memory with a capacity smaller than iconic memory but at least twice as large as working memory (Sligte, Scholte, & Lamme, 2008; Sligte et al., 2010). This memory is termed "fragile" because it is thought to be overwritten as soon as the test display is presented, hence the contrast in performance between the two conditions.

Sligte, Scholte, and Lamme (2008) and Sligte et al. (2010) claim that the contents of fragile memory are conscious despite not being directly reportable. However, as with the original Sperling paradigm, critics have argued that the improved change detection performance need not be understood in terms of a conscious comparison between the two displays but could rather be driven by unconscious perceptual representations of the specific details of the first display which influence either the appearance of, or responses to, the test display without implying rich conscious experience of the memory display (Phillips, 2011b). It is precisely this contentious issue which Vandenbroucke et al. claim to arbitrate. Their idea is to show that the memory display representations exhibit the "hallmarks of conscious perception such as perceptual organization and perceptual inference, which," they say, "are known to be absent during unconscious processing" (Vandenbroucke et al., 2012, p. 1). Their assumption then is that perceptual organization and inference require consciousness.

Using two kinds of displays which differ in terms of whether they generate Kanizsa-type modal completions, Vandenbroucke et al. show that Kanizsa displays (figure 2A) yield a distinct advantage over non-Kanizsa control displays (figure 2B) in the change detection task. This benefit was, as expected, distinctively larger in the *fragile memory* as compared to the *working memory* condition. Moreover, this distinctive added advantage for fragile memory was *not* found under isoluminant conditions under which the perceptual illusion is greatly reduced. Vandenbroucke et al. argue that together these data show that "the Kanizsa illusion was already represented in [fragile] sensory memory" (2012, p. 4). Moreover, since the illusion "is a prime example of perceptual inference, which depends on conscious perception," they conclude that representations in fragile memory must be conscious, contra the "no overflow" theorist (2012, p. 4).

How should the "no overflow" theorist respond? There is no reason, I suggest, that they should not simply accept that Vandenbroucke et al.'s data do show the existence of representations implying perceptual inference or "higher-level perceptual organization." What they should question is whether this establishes that these representations are conscious. Vandenbroucke et al. offer only one piece of evidence in support of a link between "higher-level perceptual organization" and "phenomenology," namely, that "the perception of the [Kanizsa] illusion has been shown to depend on conscious processing of its inducer elements" by Harris, Schwarzkopf, Song, Bahrami, and Rees (2011).[9] The appeal to Harris et al. (2011) is problematic in several regards, however. First, what Harris et al. show is that in a distinctive, interocular masking condition using continuous flash suppression, failure to enjoy conscious awareness of the Kanizsa inducers correlates with a failure to perceive the Kanizsa illusion.[10] However, it is doubtful that such a finding licenses the general theoretical conclusion that we *never* perceive Kanizsa-style perceptual completions in the absence of consciousness of the inducers, and in particular that one can *never* perceive such completions in relation to unattended aspects of an unmasked memory display as in Vandenbroucke et al.'s paradigm. Harris et al.'s and Vandenbroucke et al.'s paradigms are, after all, very different. Second, although Vandenbroucke et al. interpret Harris et al. as showing that perception of the Kanizsa figure "does not occur when the inducing elements are unconsciously processed" (2012, p. 6), strictly this claim goes beyond the reported data. The data show that subjects are no better in a forced-choice task when only the inducers are masked than when the whole display is masked. However, it only follows that there is no perceptual representation of the Kanizsa figure in the former case if we assume that any such representation will inevitably influence forced-choice responding. Yet that assumption contradicts evidence of high-level subliminal priming even when forced-choice responding to the prime is at chance, i.e., $d` = 0$ (see Finkbeiner, 2011 on semantic category priming, or Norman, Akins, Heywood, & Kentridge, 2014 on surface color priming).

Third, and most importantly, there is significant positive evidence that the Kanizsa illusion *can* be represented unconsciously. Vandenbroucke et al. in fact concede this, noting that "other studies [Lau & Cheung (2012) and Wang, Weng, and He (2012)] have found that the Kanizsa illusion survives crowding and breaks through interocular suppression more easily, suggesting that processing of the Kanizsa illusion can occur unconsciously or preconsciously" (2012, p. 6). In addition to the two studies just mentioned, studies of perceptual completion and illusions in neglect patients have long been claimed to reveal perceptual completion determined by unconsciously perceived inducers (e.g., Mattingley, Bradshaw, & Bradshaw, 1995; Ro & Rafal, 1996; Vuilleumier & Landis, 1998). Indeed, Block himself cites precisely such studies as "dramatic evidence" of unconscious seeing (2012, p. 12).[11] In response, Vandenbroucke et al. suggest that "the formation of the Kanizsa illusion is dependent on a diverging set of mechanisms": "basic grouping mechanisms" and "explicit figure formation," only the latter of which "is associated with conscious processing" (2012, p. 6). Say we agree. Then this simply raises the concern that the distinctive advantage in the change detection paradigm with Kanizsa figures is due to the operation of basic grouping mechanisms as opposed to explicit figure formation. Since these mechanisms are conceded to operate outside of consciousness, there can be no inference from such an advantage to phenomenal overflow.

In sum, there is no clear evidence for the claim that the representations implicated in Vandenbroucke et al.'s task must be conscious. Rather that interpretation relies directly on a tendentious assumption about the relationship between perceptual representations of certain features and consciousness which the "no overflow" theorist can reasonably reject.

6. Are Unconscious Representations Too Weak?

A distinct objection to the appeal to unconscious representations to explain findings in the Amsterdam group's change detection paradigms is that even if such representations can have the appropriate *contents*, they are simply too weak and short-lived to account for the relevant data. Recent experimental work by Sergent et al. purports to show that post-cueing attention to the location of a stimulus which has been offset 100–400 ms earlier "can produce robust enhancement of the visual experience associated with this stimulus" (2013, p. 152). This suggests that sufficiently specific unconscious representations *can* survive at least 400 ms.[12] Nonetheless, it might be maintained that unconscious representations do not persist with sufficient strength over longer timescales (i.e., ≥ 1000 ms). In support of this contention, Block (2011) cites a unique study of unconscious working memory by Soto, Mäntylä, and Silvanto (2011). This study seems to demonstrate that non-conscious representations *can* influence subject decision-making at timescales comparable to cue-delays in the kinds of change detection paradigms discussed above (i.e., ≥ 1000 ms).[13] However, since the effects are substantially smaller than the findings of the Amsterdam group, Block concludes that unconscious working memory is simply "too weak" to account for them (2011, p. 574).

However, as Block rightly notes in his discussion of the differences between the Amsterdam group's findings and those of Makovski and Jiang (2007) and Makovski, Sussman, and Jiang (2008), differences between paradigms can produce considerably different effect-sizes. Thus, Block's appeal to Soto, Mäntylä, and Silvanto (2011) is unconvincing if there are grounds for thinking that the non-conscious representations in the Amsterdam group's paradigm are likely to be significantly stronger. There manifestly are such grounds. First, in the Amsterdam group's studies, the eight-item display is presented for 250 ms (in Sligte, Scholte, & Lamme, 2008; Sligte et al., 2010) and for 500 ms (in Vandenbroucke et al., 2012), followed by a blank delay interval of at least 900 ms before either cue or test display is shown. In contrast, in place of the memory display, Soto et al. use a tilted Gabor grating initially displayed for just 16.67 ms and then immediately (and strongly) masked. Thus, the stimuli are markedly different in nature, presentation time, and context. Accordingly, the strength of the non-conscious representations evoked should be expected to be substantially different. Second, in Soto et al.'s task, subjects report *no* conscious awareness of the Gabor patch at all (in relevant trials). In contrast, in the Amsterdam group's task, subjects are perfectly aware that there are items or figures in all eight relevant locations in the memory display—the "no overflow" theorist's claim is only that subjects do not perceive specific features of such items (e.g., orientation or configuration). Since it is plausible that non-conscious representations of stimulus features will be weaker when subjects do not report awareness of the stimulus at all, as opposed to when they report seeing it though not its determinate characteristics, there is again good reason to think that the unconscious representations in Soto et al.'s paradigm will be weaker.

Given these differences, Soto et al.'s findings in no way render empirically implausible a "no overflow" interpretation of the Amsterdam group's findings in terms of non-conscious representations. Moreover, since these findings are the only ground which Block adduces to reject long-lived unconscious "working memory" representations, the case against such representations lapses.

7. Back to Naïve Introspection

As a final objection to the "no overflow" theorist's "appeal to unconscious iconic memory that is specific enough to provide the information to do the task," Block claims that such a supposition "goes counter to what subjects report about their own experience" (2011, p. 570). To support this, he quotes

Baars: "Subjects—and experimenters serving as subjects—continue to insist that they are momentarily conscious of all the elements in the array" (1988, p. 15). I and various others have criticized this brute appeal to introspection as inadequate elsewhere (e.g., Byrne, Hilbert, & Siegel, 2007; Papineau, 2007; Phillips, 2011b; Sergent & Rees, 2007). Part of the problem is that momentary consciousness of all the elements is consistent with a "no overflow" account so long as subjects aren't momentarily conscious of each individual element as a *specifically* identified letter (as opposed to some letter-like form or other). And it is not clear that subjective reports really do distinguish between these two possibilities.

Inadequate or not, what the above discussion illustrates in various ways is that it is extremely hard to see how to move beyond such appeals to introspection in debates about the nature of experience. To do so seems inevitably to involve us in substantial and contentious assumptions aimed at bridging the gap between empirical findings implicating certain kinds of representation and claims about phenomenal experience (compare Martin, 1997). The repeated lesson of this paper is that providing empirical evidence for the richness or poverty of experience is constantly thwarted by the necessity of appealing to such assumptions. As a result, at least in our present state of understanding, there seems little else for Block and fellow overflow theorists to do but insist that "no overflow" accounts fail to do justice to our naïve phenomenology. There is likewise little else for opponents of overflow to do but insist that Block and others have got the phenomenology wrong.

Notes

1. A related view, defended by de Gardelle, Sackur, and Kouider (2009) and Kouider et al. (2010), is that subjects enjoy different levels of conscious access to the features in the uncued rows, a low-level form of partial access to scattered fragments/features of the letters and a higher-level form of access to the letters simply as generic letter forms. For critical discussion of both views see Block (2011).
2. Compare Coltheart (1980) on the distinction between informational and phenomenological persistence.
3. It is rather less clear how to adapt the partial awareness hypothesis of de Gardelle, Sackur, and Kouider (2009) and Kouider et al. (2010) to this context given that fragments of a cyan-B, orange-K, and mauve-T will themselves presumably be cyan, orange, and mauve. As a result, some alternative understanding of fragmentary awareness will be needed here.
4. Bronfman et al. do convincingly demonstrate that color is represented outside of focal attention contra Lau and Rosenthal (2011). But lack of color experience outside attention is not an essential commitment of a "no overflow" theory.
5. Note that despite the standard invocation of color to introduce qualia or phenomenal consciousness (Block, 2014, p. 445; Bronfman et al., 2014, p. 1402, footnote 1; Tye, 2013), Bronfman et al. do not suppose that colors are *essentially* phenomenal in that they could not be perceptually represented in the absence of consciousness. This claim would of course block any "no overflow" interpretation. However, such a claim would need to confront recent evidence that color *can* be perceptually represented outside of consciousness, even when the latter is measured objectively. In particular, using a metacontrast masking paradigm, Norman et al. (2014) found priming effects implicating surface color representations (implying constancy based perception) despite $d`$ not differing significantly from zero. In Phillips (forthcoming-a), I discuss whether such cases of perceptual representation amount to perception of the same fundamental kind as ordinary conscious perception by the individual. However, issues of attributability to the individual are orthogonal to the issues of present concern.
6. One might try to combine naïve realism with the claim espoused by Gillett and Rives that "the world contains absolute determinate properties, but no determinable properties" (2005, p. 501; see also Armstrong, 1978, chapter 22), to yield the view that perceptual experience could only present determinate properties. However, defenders of overflow such as Block are hardly likely to rest their case on such foundations.
7. For another everyday example, one might also point to letters seen at a distance, or with uncorrected short-sight (e.g., Tye, 2003).
8. In addition to Bronfman et al.'s own paper, there is also evidence of ensemble representations of average color (Maule, Witzel, & Franklin, 2014). Such data appears also to reveal "a mechanism sensitive to the variance [i.e. color diversity] in hue present in a scene" (Maule, Witzel, & Franklin, 2014, p. A100). Earlier work on color contrast and constancy also suggests the operation of such a mechanism (see Brown & MacLeod, 1997). However, none of this work settles whether individual colors are *consciously* represented outside of focal attention.
9. They also write that when representations go beyond mere physical input and add "perceptual qualities" that we have conscious representation. In the absence of further argument, I see no reason why the "no overflow" theorist should not simply deny this claim. Here compare Burge's claim that "consciousness is constitutively

neither necessary nor sufficient for perception" (2010, p. 305). They might also question whether perceptual completion suffices for genuine *perceptual* representation (compare Burge, 2010, p. 418).

10. I bracket here important concerns one might have about the subjective measure of awareness used in this and similar studies (compare Phillips forthcoming-a, forthcoming-b).
11. As I argue in Phillips (forthcoming-b), whether this is the right interpretation of these studies depends upon vexed issues to do with our measure of consciousness. The point here is only that the existence of cases in which we have unconscious representations of inducers yielding perceptual completions is a reasonable empirical hypothesis even by the lights of overflow theorists.
12. Does this work decisively support a "no overflow" interpretation of Sperling along the lines of Phillips (2011b)? It does not. Overflow theorists can interpret the post-cue not as enhancing visual experience, but rather as improving attention-based retention and subsequent access to already rich experience. Once again, the gap between stories told purely in terms of representations and claims about phenomenology yawns wide.
13. Here, following Block, I accept that the relevant representations are unconscious. However, Soto et al.'s study principally uses a biased measure to assess awareness and so could certainly be challenged on this score (Phillips, forthcoming-a, forthcoming-b).

Acknowledgements

I am very grateful to the Editor and to two anonymous reviewers for this journal for their comments on an earlier version of this material which led to very substantial improvements. This work was supported by the Leverhulme Trust under grant number RF-2013-278.

References

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science, 19*, 392–398.
Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science, 12*, 157–162.
Armstrong, D. M. (1978). *A theory of universals*. Cambridge: Cambridge University Press.
Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.
Block, N. (1983). The photographic fallacy in the debate about mental imagery. *Noûs, 17*, 651–661.
Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences, 18*, 227–287.
Block, N. (2007). Consciousness, accessibility and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences, 30*, 481–548.
Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences, 15*, 567–575.
Block, N. (2014). Rich conscious perception outside focal attention. *Trends in Cognitive Sciences, 18*, 445–447.
Block, N., Carmel, D., Fleming, S. M., Kentridge, R. W., Koch, C., Lamme, V. A. F., Lau, H., & Rosenthal, D. (2014). Consciousness science: Real progress and lingering misconceptions. *Trends in Cognitive Sciences, 18*, 556–557.
Bronfman, Z., Brezis, N., Jacobson, H., & Usher, M. (2014). We see more than we can report: 'Cost free' color phenomenality outside focal attention. *Psychological Science, 25*, 1394–1403.
Brown, R. O., & MacLeod, D. I. A. (1997). Color appearance depends on the variance of surround colors. *Current Biology, 7*, 844–849.
Burge, T. (2007). Psychology supports independence of phenomenal consciousness. *Behavioral and Brain Sciences, 30*, 500–501.
Burge, T. (2010). *The origins of objectivity*. Oxford: Oxford University Press.
Byrne, A., Hilbert, D. R., & Siegel, S. (2007). Do we see more than we can access? *Behavioral and Brain Sciences, 30*, 501–502.
Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research, 43*, 393–404.
Chubb, C., Nam, J.-H., Bindman, D. R., & Sperling, G. (2007). The three dimensions of human visual sensitivity to first-order contrast statistics. *Vision Research, 47*, 2237–2248.
Cohen, M., & Dennett, D. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences, 15*, 358–364.
Coltheart, M. (1980). Iconic memory and visible persistence. *Perception and Psychophysics, 27*, 183–228.
de Gardelle, V., Sackur, J., & Kouider, S. (2009). Perceptual illusions in brief visual presentations. *Consciousness & Cognition, 18*, 569–577.
Dretske, F. (2006). Perception without awareness. In T. S. Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 147–180). Oxford: Oxford University Press.
Finkbeiner, M. (2011). Subliminal priming with nearly perfect performance in the prime-classification task. *Attention, Perception, and Psychophysics, 73*, 1255–1265.

Gillett, C., & Rives, B. (2005). The non-existence of determinables: Or, a world of absolute determinates as default hypothesis. *Noûs, 39*, 483–504.
Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention Perception & Psychophysics, 72*, 1825–1838.
Harris, J. J., Schwarzkopf, D. S., Song, C., Bahrami, B., & Rees, G. (2011). Contextual illusions reveal the limit of unconscious visual processing. *Psychological Science, 22*, 399–405.
Kouider, S., de Gardelle, V., Sackur, J., & Dupoux, E. (2010). How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Sciences, 14*, 301–307.
Kulvicki, J. (2014). *Images*. London: Routledge.
Landman, R., Spekreijse, H., & Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research, 43*, 149–164.
Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences, 15*, 365–373.
Lau, J. S. F., & Cheung, S. H. (2012). Illusory contour formation survives crowding. *Journal of Vision*, 12(6), 15, 1–12.
Mack, A., & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.
Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide*. London: Erlbaum.
Makovski, T., & Jiang, Y. V. (2007). Distributing versus focusing attention in visual short-term memory. *Psychonomic Bulletin & Review, 14*, 1072–1078.
Makovski, T., Sussman, R., & Jiang, Y. V. (2008). Orienting attention in visual working memory reduces interference from memory probes. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 34*, 369–380.
Martin, M. G. F. (1997). The shallows of the mind. *Proceedings of the Aristotelian Society, 74*, 75–98.
Mattingley, J. B., Bradshaw, J. L., & Bradshaw, J. A. (1995). The effects of unilateral visuospatial neglect on perception of Müller-Lyer illusory figures. *Perception, 24*, 415–433.
Maule, J., Witzel, C., & Franklin, A. (2014). Getting the gist of multiple hues: Metric and categorical effects on ensemble perception of hue. *Journal of the Optical Society of America A, 31*, A93–A102.
McConkie, G. W., & Rayner, K. (1975). The span of effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*, 578–586.
Norman, L. J., Akins, A., Heywood, C. A., & Kentridge, R. W. (2014). Color constancy for an unseen surface. *Current Biology, 24*, 2822–2826.
Papineau, D. (2007). Reuniting (scene) phenomenology with (scene) access. *Behavioral and Brain Sciences, 30*, 521.
Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*, 739–744.
Phillips, I. B. (2011a). Attention and iconic memory. In C. Mole, D. Smithies, & W. Wu (Eds.), *Attention: Philosophical and psychological essays* (pp. 204–227). Oxford: Oxford University Press.
Phillips, I. B. (2011b). Perception and iconic memory. *Mind & Language, 26*, 381–411.
Phillips, I. B. (forthcoming-a). Unconscious perception reconsidered.
Phillips, I. B. (forthcoming-b). Consciousness and criterion: On Block's case for unconscious seeing. *Philosophy & Phenomenological Research*.
Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7*, 65–81.
Reingold, E. M., & Merikle, P. M. (1990). On the inter-relatedness of theory and measurement in the study of unconscious processes. *Mind & Language, 5*, 9–28.
Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8*, 368–373.
Ro, T., & Rafal, R. D. (1996). Perception of geometric illusions in hemispatial neglect. *Neuropsychologia, 34*, 973–78.
Sergent, C. & Rees, G. (2007). Conscious access overflows overt report. *Behavioral and Brain Sciences, 30*, 523–524.
Sergent, C., Wyart, V., Babo-Rebelo, M., Cohen, L., Naccache, L., & Tallon-Baudry, C. (2013). Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Current Biology, 23*, 150–155.
Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE, 3*, e1699.
Sligte, I. G., Vandenbroucke, A. R. E., Scholte, H. S., & Lamme, V. A. F. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in Psychology, 1*(175), 1–10.
Soto, D., Mäntylä, T., & Silvanto, J. (2011). Working memory without consciousness. *Current Biology, 21*, R912–913.
Sperling, G. (1960). The Information Available in Brief Visual Presentations. *Psychological Monographs, 74*, 1–29.
Stazicker, J. (2011). Attention, visual consciousness and indeterminacy. *Mind & Language, 26*, 156–184.
Tye, M. (2003). Blurry image, double vision and other oddities: New problems for representationalism? In Q. Smith & A. Jokic (Eds.), *Consciousness: New philosophical perspectives* (pp. 7–32). Oxford: Oxford University Press.
Tye, M. (2006). Content, richness, and fineness of grain. In T. S. Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 504–530). Oxford: Oxford University Press.
Tye, M. (2013). Qualia. In E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2013 ed.). Retrieved from http://plato.stanford.edu/archives/fall2013/entries/qualia/

Vandenbroucke, A. R. E., Fahrenfort, J. J., Sligte, I. G., & Lamme, V. A. F. (2014). Seeing without knowing: Neural signatures of perceptual inference in the absence of report. *Journal of Cognitive Neuroscience, 26*, 955–69.
Vandenbroucke, A. R. E., Sligte, I. G., Fahrenfort, J. J., Ambroziak, K. B., & Lamme, V. A. F. (2012). Non-attended representations are perceptual rather than unconscious in nature. *PLoS ONE, 7*, e50042.
Vuilleumier, P., & Landis, T. (1998). Illusory contours and spatial neglect. *Neuroreport, 9*, 2481–2484.
Wang, L., Weng, X., & He, S. (2012). Perceptual grouping without awareness: superiority of Kanizsa triangle in breaking interocular suppression. *PLoS ONE, 7*, e40106.
Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research, 24*, 55–62.