DOI: 10.1111/mila.12217

WILEY

ORIGINAL ARTICLE

Impoverished or rich consciousness outside attentional focus: Recent data tip the balance for *Overflow*

Zohar Z. Bronfman[1,2] | Hilla Jacobson[3] | Marius Usher[1,4]

[1]School of Psychology, Tel-Aviv University, Tel-Aviv, Israel

[2]The Cohn Institute for the History and Philosophy of Science and Ideas, Tel-Aviv University, Tel-Aviv, Israel

[3]Departments of Philosophy and Cognitive Science, The Hebrew University of Jerusalem, Jerusalem, Israel

[4]Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv, Israel

Correspondence
Zohar Z. Bronfman, Department of Psychology, Sharet Building, Tel-Aviv University, Tel-Aviv, Israel.
Email: zoharbronfman@gmail.com

Funding information
Israeli Science Foundation, Grant/Award Number: 1413/17

The question of whether conscious experience is restricted by cognitive access and exhausted by report, or whether it *overflows* it—comprising more information than can be reported—is hotly debated. Recently, we provided evidence in favor of Overflow, showing that observers discriminated the color-diversity (CD) of letters in an array, while their working-memory and attention were dedicated to encoding and reporting a set of cued letters. An alternative interpretation is that CD-discriminations do not entail conscious experience of the underlying colors. Here we argue, based on conceptual considerations and consistency with neuroscience and phenomenology, in favor of the Overflow interpretation.

KEYWORDS

attention, neural correlates of consciousness, phenomenal/access consciousness, summary statistics, working-memory

1 | INTRODUCTION

An enduring debate on the nature of visual consciousness pertains to whether it is subject to a limited access capacity, reflecting an attentional bottleneck (about three to four items: Sperling, 1960; Luck & Vogel, 1997), or whether it is “richer” and *overflows* cognitive access. The latter position was eloquently advocated by Ned Block (Block, 1995, 2007, 2008, 2011), who argued for a distinction between phenomenal consciousness (how having an experience feels)—which is rich—and access consciousness (characterized by report and cognitive availability: Baars, 1993; Dehaene, Changeux, Naccache, Sackur & Sergent, 2006)—which is capacity limited. The Overflow position has attracted strong criticism within both the cognitive sciences and the philosophy of mind (e.g., Cohen & Dennett, 2011; Dehaene, 2014; Kouider, Sackur & De Gardelle, 2012). Overflow opponents have

wileyonlinelibrary.com/journal/mila

argued that the assumption of a nonaccessible component of (phenomenal) consciousness, in addition to access consciousness and to unconscious processes, is ill-motivated and redundant. They consequently proposed alternative interpretations of the same evidence that Block relied on to support his position (see below), which do not rely on phenomenal overflow. According to such theorists, consciousness outside focal attention is sparse and impoverished (Cohen & Dennett, 2011; Cohen, Dennett & Kanwisher, 2016; Kouider et al., 2012; Lau & Rosenthal, 2011; Noë & O'Regan, 2000) or indeterminate (Phillips, 2011; Stazicker, 2011).

In a recent experimental study we have reported results that we believe provide relevant new evidence in favor of the Overflow position (Bronfman, Brezis, Jacobson & Usher, 2014; see also discussion in Block, 2014). These results, however, have provoked a vigorous opposition from no-Overflow theorists (Cohen et al., 2016; Gross & Flombaum, 2017; Phillips, 2016; Richards, 2015). In particular, in a recent article, Phillips (2016) has disputed our interpretation of the results, stating that the conclusions are based on poorly motivated auxiliary assumptions. According to Phillips, counter interpretations, similar to those that have been employed to block the arguments for Overflow in previous experiments (e.g., Sperling's), can also be used to undercut our Overflow argument, and therefore the two competing positions are left in a dialectical tie. The main goal of the present paper is to augment the case for a weak version of Overflow, according to which observers have visual awareness of at least *some* visual properties that are outside focal attention, primarily by appealing to our recent results. The case we shall present in favor of Overflow will include answers to the substantial challenges set up by Phillips. We will do so by discussing the counter-interpretations, analyzing their commitments, and providing some of the motivations that were missing in our original work. We start with a review of the Overflow debate and of the previous experimental evidence relevant to it, including our recent work (Bronfman et al., 2014). We then discuss the counter arguments (Fink, 2015; Gross & Flombaum, 2017; Phillips, 2011, 2016; Richards, 2015; Stazicker, 2011) and clarify their assumptions. Finally, we argue that theoretical-phenomenological considerations, as well as principles and data from cognitive science and neuroscience, tip the balance in favor of the Overflow.

2 | REVIEW OF THE OVERFLOW DEBATE

Block has based his case for the Overflow position on the seminal experiments by George Sperling on iconic memory (Sperling, 1960), as well as on a number of more modern follow ups (Landman, Spekreijse & Lamme, 2003; Sligte, Scholte & Lamme, 2008). In Sperling's experiments, the subjects are briefly (for 100–200 ms) presented with an array of (3 × 4) letters for report. When the report is free, the subjects can only report about three to four letters, although they typically also claim that they saw all letters but "lost" them before they could report them. This result is thought to reflect the capacity limitation of attentional access that limits encoding into an enduring working memory (Sperling, 1960). The surprising result is that if a cue is presented after the array has disappeared (but no later than 500 ms), instructing subjects to report the letters in one of the three rows, the subjects are able to report (almost) all the letters of that row. The conclusion that was suggested by Sperling, and is widely accepted within cognitive sciences, is that high-resolution information about the letters is maintained in *iconic* memory, which is fragile and decays within about 500 ms, but allows transfer of some of the information to a durable working memory that allows report, if attention is directed to it before it decays. The conscious status of the representations in iconic memory (or in a similarly high capacity memory system, termed "fragile STM"; Sligte et al., 2008), is what stands at the focus of the debate between the Overflow and the no-Overflow theorists.

While Sperling interpreted his results to indicate that subjects typically enjoy a rich conscious experience of the array,[1] which decays within about 500 ms (see also Block, 1995, 2007; Dretske, 2006; Tye, 2006), proponents of the no-Overflow view have suggested an alternative account. Accordingly, iconic memory consists of unconscious representations, and it is only upon the presentation of the cue that the content of the cued elements is rendered conscious, by the deployment of visual attention (Phillips, 2011; Stazicker, 2011). While this interpretation is consistent with postdiction experiments (e.g., Eagleman & Sejnowski, 2000, in which cues were shown to affect perception of events half a second before they appeared; Phillips, 2011) and has been supported on the grounds of appearing more parsimonious (Cohen & Dennett, 2011; Kouider et al., 2012; but see Block, 2012), it is at odds with subjects' report that they saw more than they could subsequently report. To account for these introspective reports, the no-Overflow theorists have appealed to a distinction between: (a) generic (indeterminate, undetailed) representations—e.g., the representation of an item as having a “letter-like form”; and (b) specific (determinate, detailed) representations (“X”, “L”, etc.). Their proposal is that the introspective reports “I saw more”, are based on what Fink (2015; see also Block, 2015) has labeled “solely generic phenomenology”—i.e., phenomenology that is generic and does not include “an accompanying and subsumable concretum” (Fink, 2015, p. 8). Accordingly, before the cue appears, only the three to four letters to which the subjects attend are represented in a conscious specific manner, while other letters are represented in a solely generic or indeterminate manner (which specifies only that “there are some letter-shaped forms”). As recently argued by Phillips (2016), this appears to leave the debate undecided, as both theories can equally explain the basic results.

In what follows we discuss two ways to defend a type of Overflow theory. To clarify, the type of Overflow we wish to defend here does not involve a distinction between phenomenal and *access* in the weak sense of being *accessible* in principle, but rather in the strong sense of being accessed in the actual case under some specific empirical conditions. Second, we limit *access* to an even more restrictive type of *robust* access (to working memory and thus to report). The latter qualification means that we do not consider contents that are transiently accessed (in a fragile way that does not allow report) to be robustly accessed and we argue that one can have phenomenal experiences without robust access, that is, that we can experience more than we can report. While this position is more modest, it is nevertheless contrary to some influential conceptualizations of consciousness in terms of access to global workspace or working memory (Baars, 1993; Dehaene et al., 2006; Kouider et al., 2012). Finally, we will limit ourselves to arguing for the possibility of phenomenal experience of visual elements (e.g., rectangles or colors), rather than of composites (e.g., objects or letters), in the absence of access to focal attention, working memory and report.

Sections 2 and 4.1 challenge the plausibility of the solely generic phenomenology states for some particular situations. Section 4.2 attacks the association of the specific (high resolution) representations with unconscious processes, by examining the debate in a wider context of theory and data from neuroscience and cognitive science. This section will be mostly based on data from our recent experiment (to be presented in section 3). We will argue that these considerations favor the Overflow account.

2.1 | The Landman experiment

We start with an experiment carried out by Landman et al. (2003), which combines the Sperling paradigm with a change blindness design (Rensink, O'Regan & Clark, 1997), and which challenged the

[1] Sperling (1960) used the terms “visual image” and “persistent sensation.”

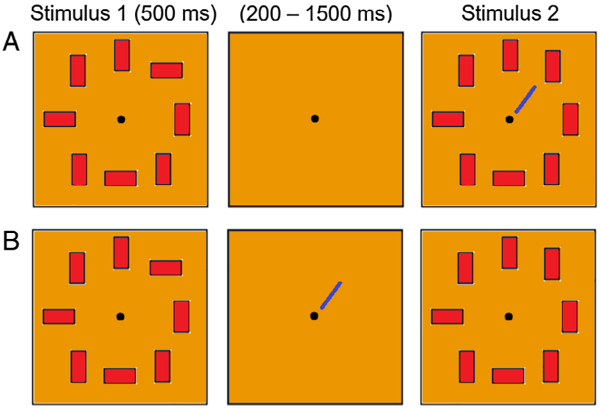

FIGURE 1 Illustration of the Landman experiment; based with modifications on Landman 2003. The task here is to detect if the cued rectangle has changed in orientation or not; the critical difference is whether the cue is presented before the second array is shown (i.e., before the trace is erased; bottom panel) or together with it (top panel). Top: the late cue condition; bottom: the early cue condition [Color figure can be viewed at wileyonlinelibrary.com]

coherence of the solely generic phenomenology account (Block, 2011). In this experiment, the Sperling letter-array is replaced by a circular arrangement of eight rectangles, which appears for several hundred milliseconds, followed by another array, which is either identical or has one rectangle changed in orientation (see Figure 1).[2] The observers are asked if the cued rectangle has changed. As in the original Sperling experiment, the observers can only report accurately on about three to four of the eight rectangles, if the cue is presented simultaneously with the second display, and hence erases the original trace from iconic memory. On the other hand, if the cue is presented during a blank interval (about 1.5 s after the rectangle stimuli have disappeared) and before the second array appears, the observers' capacity to detect the change increases to about six to eight elements (Landman et al., 2003). Thus, as in Sperling's experiment, it appears that the observers can maintain the representations of roughly eight rectangles in an iconic memory store (or fragile short-term memory; Sligte et al., 2008), from which they can access only a subset, by encoding them into a durable working memory trace. Also, as in Sperling's experiment, the observers reported that they saw all rectangles, although they could only notice differences in a few of them.

Note, first, that in order to account for this result, the no-Overflow theorist has to admit that all rectangles are represented with enough specificity to distinguish horizontal from vertical orientations. To deny phenomenal awareness for unattended items, however, she must assume that: (a) up to the cue presentation (bottom panel of Figure 1), the specific/determinate representations of the rectangles at unattended locations (i.e., at least half of the eight) are unconscious; (b) in trials in which an unattended element is probed, the presentation of the cue transforms the specific unconscious representation into a conscious one (note that in some versions of this experiment the cue is presented at delays of 1.5 and up to 4 s after the stimulus disappeared, which is longer than the range of postdiction effects). Critically, in order to account for the introspective report ("I saw all the rectangles"), the no-Overflow theorist must assume that in addition to the specific unconscious representations of unattended rectangles and to the specific conscious representations of the attended ones, there is also a

[2] A similar experiment, which was carried out with more complex shapes, rather than with rectangles, showed similar results (Sligte, Vandenbroucke, Scholte & Lamme, 2010). For simplicity we focus here on the original one.

solely generic conscious representation (of the unattended rectangles), which is utterly silent about their orientation.

2.2 | The coherence of the solely generic conscious state for rectangles

As argued by Block (2011), the Landman experiment provides us with a stringent case study of the solely generic phenomenal component in the no-Overflow account. While, for the original Sperling experiment, solely generic phenomenology may plausibly correspond to blurred (or fragmented; Kouider et al., 2012) letter-like shapes, we can now ask what does it mean to have a generic phenomenal representation of a rectangle, and in particular, whether it is plausible or even coherent to be conscious of seeing a (generic) rectangle without also being conscious of its (specific) orientation. Phillips (2016) argues against Block's negative answer to the question above, by appealing to the photographic fallacy (Block, 1983). According to the photographic fallacy, it is a mistake to assume that when one is conscious of a visual object, all of its properties must also be represented in a determinate way. Interestingly, however, Block's photographic fallacy paper, was aimed to show how depictive (but not naïve or fully realistic photographical) representations can withstand challenges of indeterminacy (e.g., the number of stripes in a mental image of a tiger[3]), and not to argue that depictive representations are inadequate to imagery or conscious experience. In his reply to Fink, Block (2015) expresses doubt that there can be "generic phenomenology of an oriented rectangle that does not specify the rough orientation of the rectangle" (p. 4). In this section, we aim to further support this position, before we will examine (sections 3 and 4) a new type of data that makes the solely generic phenomenology in Sperling-type experiments, even less plausible.

One way to understand solely generic phenomenology is in relation to the determinate-determinable model (Stazicker, 2011). According to this model:

> To represent something indeterminately… is to represent it as instantiating a determinable property, without commitment as to which determination of that determinable it instantiates. Roughly, property A determines property B where to have A is to have B in a specific way. For example, the property of being crimson (A) or the property of being scarlet (A′), are each determinates of the property of being red (B) (Stazicker, 2011, p. 170).

Following Stazicker (2011), let us consider what an indeterminate phenomenal representation of a rectangle—one with no orientation—may consist of, or to put this in experiential terms, how does such a rectangle look like? If we grant, for present purposes, Stazicker's considerations about perception-without-attention having a lower spatial resolution, this would suggest that we should see a blurry rectangle (say, one that is convolved with a Gaussian; see Figure 2). But now we have two options. Either the blur is strong (broad Gaussian, Figure 2, right panels), in which case we do not see the orientation but neither do we see the rectangle, or it is minor enough (narrow Gaussian, Figure 2, left panels), in which case we see the orientation, as well as the rectangularity.

This line of reasoning suggests that there is no way to visualize a (nonsquare) rectangle without also visualizing it with a specific orientation. We contend that this is because orientation is a constitutive property of any experienced particular rectangle, whose properties (aspect-ratio, orientation and magnitude) are determined (up to certain precision). According to the Overflow position that we endorse, one can visually experience a shape with a reduced precision on a spatial constitutive

[3] Block (1983) suggests that mental images represent like drawings, which have special structure, but omit details. He also discussed a suggestion by Fodor (1975) that they represent like blurred photographs (Figure 2).

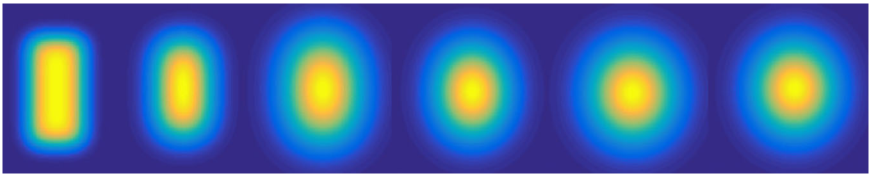

FIGURE 2 A sequence of seven rectangles, which are convolved with Gaussians of increasing radia (from left to right). As the Gaussian radius increases, high spatial frequency details are lost, resulting in the rectangle appearing increasingly more blurred. At the low-blur we experience this as a vertical rectangle, while at the other extreme as a circular spot. There is no place on this continuum, where we experience a rectangle without a specific orientation [Color figure can be viewed at wileyonlinelibrary.com]

property, such as orientation, but not with *null* precision on *all* visual properties, which would be the case for a totally orientationless rectangle. This raises the question of what visual properties are. Is "being a rectangle" a visual property satisfying the criterion above, in the absence of any spatial-orientation information? Based on psychophysics research and visual neuroscience, we contend that visual properties are based on a set of primary visual features (e.g., orientated lines and gratings, their spatial binding and colors, etc.), which satisfy some metric relations on an analog continuum (red is closer to orange than to yellow; | is closer to \ than to –), and which are subject to experimental tests.[4] The property of "being a rectangle," on the other hand, satisfies no such metric properties, as it involves an infinite set of visual objects with different visual properties.

We believe that the idea of visually experiencing a rectangle without any orientation (in contrast to the idea of having a linguistic/symbolic representation with the content of "being a rectangle") is a challenging one. This is because an experience of "looking like a rectangle" without the possession of a characteristic visual property such as its orientation (even if imprecise), appears to involve a radical re-interpretation of the term "looking like." Note that we by no means dispute here that one can entertain a propositional thought of "looking like a rectangle" and even experience some kind of cognitive phenomenology associated with it (such as feeling confident that the proposition is correct). Rather, we contend that as soon as visual phenomenology appears, *some* concrete properties (within some, even if very coarse-grained precision range) need to be a part of the experience. However, no-overflow proponents may insist that there can be nonlinguistic visual experiences, which have nevertheless totally undetermined visual properties. Perhaps, for example, people can experience a rectangle as being, disjunctively, "vertical or horizontal." While this would match the introspective report of subjects in the Landman experiment of seeing rectangles that are either vertical or horizontal, it seems to contradict basic facts about the nature of visual experience, such as its abhorrence of disjunctive states, which is well illustrated by a variety of rivalry phenomena (Leopold & Logothetis, 1996; when forced to experience ambiguous disjunctive experiences, the visual system tends to oscillate between the pure components, rather than meshing them.

A different way of understanding the solely generic phenomenology for rectangles corresponds to the idea that a generic conscious representation is akin to a sort of linguistic description rather than an image-like representation. If this idea is accepted, then the test of whether it is possible to depict or visualize a percept is simply irrelevant—the supposition that it is a constraint on visual contents that they can be depicted, in any way, should be rejected. Specifically, one may suggest that the solely generic phenomenology of seeing rectangles in the Landman experiment during the empty gap

[4] For example pop-out. A vertical lines pops out among horizontal ones, and a red patch among blue ones; no such pop-out takes place for triangles among other nontriangle type polygons, which are heterogeneous with regards to primary visual properties such as orientation or size.

before the cue presentation is grounded in symbolic (rule-based or propositional) representations (e.g., Fink, 2015). Under this interpretation, the solely generic phenomenology in the Landman experiment (Figure 1, bottom, before the cue) corresponds to something like entertaining the proposition: there are a number of rectangles in front of me. A long debate within psychology has involved the question of whether visual imagery (which we can take to involve visual phenomenology) operates on visual image-like representations (Kosslyn, Thompson & Ganis, 2006) or on purely propositional ones (Pylyshyn, 1973, 2003). Note, however, that the imagery debate was not about the nature of conscious visual experience. As Pylyshyn (1973) states:

> Imagery is a pervasive form of experience and is clearly of utmost importance to humans. We cannot speak of consciousness without, at the same time, implicating the existence of *images.* The main question that is raised is whether the concept of *image* can be used as a primitive in psychological theories. And finally, must *images always be conscious*? (Pylyshyn, 1973, p. 2, italics added).

Thus, the contention point was not the existence of conscious images in visual experience but rather whether such conscious images are being deployed in the typical "mental imagery" experiments, which, alternatively may be explained by the deployment of (unconscious) propositional representations. A large set of ingenious experiments were able to demonstrate that visual imagery processes have many of the signatures predicted by analog image theories, such as: i) sensitivity to details that are size-dependent (Kosslyn, 1975), and ii) they can be integrated with percepts to create a single composite representation (Brockmole, Wang & Irwin, 2002; Lewis, Borst & Kosslyn, 2011).[5] Note, however, that Pylyshyn never argued that the propositional representations are the bearers of visual consciousness.

We can see only one no-Overflow account for the Landman experiment, which does not give up on the notion of visual experience and is consistent with data in visual science. This account admits the experience of rectangles with orientation specificity during the gap, but insists that in the absence of focal attention, the binding between the locations and the orientations is random, as suggested by the presence of illusory conjunctions in brief displays without focal attention (Treisman & Schmidt, 1982). Note that this requires extending the range of illusory conjunctions from the domain reported (color and shape) to location and shape. This position is also equivalent to a recent proposal that generic phenomenal states involve summary statistics (Cohen et al., 2016; in this case the summary involves: fraction-X of vertical rectangles, fraction 1-X or horizontal rectangles, with no binding to location). A similar position was made by some no-Overflow opponents in the Sperling experiments, by suggesting that what people perceive outside attentional focus are letter-fragments (De Gardelle, Sackur & Kouider, 2009; Kouider & Dehaene, 2007). Accordingly, the role of the cue is to resolve the binding (between letter fragments in the Sperling experiment, or between orientations and locations, in the Landman one).

While we do not endorse this position, we do not aim to dispute it here. Rather, we wish to support a mild version of the Overflow account, according to which observers have phenomenal experience of visual *elements* outside focal attention, which they cannot access, and to which they have a transient and fragile access—one that is too fleeting to allow report. To demonstrate this we turn to a novel paradigm, which does not probe the access of the elements via their binding with specific locations, and in which the elements are not composed of perceptual sub-units. Furthermore, these results

[5] In Brockmole et al., it is demonstrated that although observers cannot integrate two visual percepts (corresponding to visual arrays of black and white squares), which are separated by an 100 ms empty gap, they can integrate the second percept with the first, if they have enough time (about 1.3 s) to form a visual-image of the first before the second array is presented.

will serve to mount a further attack on the no-Overflow view, by uncovering and criticizing the association between the precise-representations of unattended elements and unconscious processes to which this view is committed.

3 | EXPERIENCING THE COLOR-DIVERSITY OF UNATTENDED LETTERS IN A SPERLING TYPE TASK

In a recent paper we have attempted to demonstrate that when carrying out a Sperling-type task, observers have awareness of some visual properties of some unattended letters. To do this we modified the Sperling experiment in the following way (Bronfman et al., 2014). First, we presented a precue before (rather than after) the onset of the display. The cue indicated the row from which the letters should be memorized (transferred to working memory) for future recall. This procedure ensured that observers focus their attentional processes on a specific row, and thus that the rest of the array was not subject to focal attention.[6] Second, we presented the letters in colors, which were generated so as to create, on different (randomized) trials, either high or low color-diversity (CD) (see Figure 3).

While the primary task, which the participants performed in all the experimental blocks, and for which they received feedback and reward, was to report a cued letter from within the precued row (see Figure 3c), on some trials they were also queried about the colors of either the precued (attended) *or* the noncued (unattended) letters (this was done exclusively, on different blocks). In particular, observers were required to indicate whether the CD of the unattended letters was high or low (see Figure 3a,b, for a description of how the colors were generated, so as to create high/low CD, in either the cued-row or in the noncued row, independently).[7]

The experimental results showed that observers were able to detect correctly the CD of the unattended letters (Experiment 1: 67%; Experiment 3, which involved a stricter manipulation of CD, 62%; chance level: 50%), although they did not appear to divert attentional resources from the letter-report task to the color task. There was no difference in letter recall accuracy between experimental blocks that only tested the primary task of letter recall and blocks in which the CD task was queried and the performance was far from ceiling and close to the regular estimate (about three items) that is obtained in such tasks. Our results thus demonstrate that while observers are focusing their attentional resources on a row of letters for encoding them into WM they are also able to report an important high-order statistical property of the colors of the unattended letters. We believe that there is no controversy about this conclusion (Gross & Flombaum, 2017; Phillips, 2016; Richards, 2015; Ward, Bear & Scholl, 2016), which was recently replicated in another laboratory (Ward et al., 2016). There are, however, divergent interpretations on whether seeing the individual colors outside the focus of attention is necessary for making CD judgments and, if they are, how many such colors one has to see.

First, in a recent paper Gross and Flombaum (2017) argue that to account for the level of CD detection (in the range manipulation version of the Bronfman et al., 2014 experiments) it would suffice to discriminate three colors at unattended rows, and then make a decision based on the largest distance between them on the color wheel (larger/smaller than six), which was the range of colors for low CD arrays in the range-version of the CD experiments. Note that this estimation relies on two crucial assumptions. First, that these three colors are perceived without any imprecision—i.e., that

[6] The rest of the array may benefit from residual diffuse attention sources, but those are also present in the Sperling experiment, and they do not enable cognitive access to an enduring WM store.

[7] In Experiment 3 we carried out a more strict manipulation of CD, in which the color s were always sampled from the full range, and thus there is no confound of average color (Bronfman et al., 2014, Supplement).

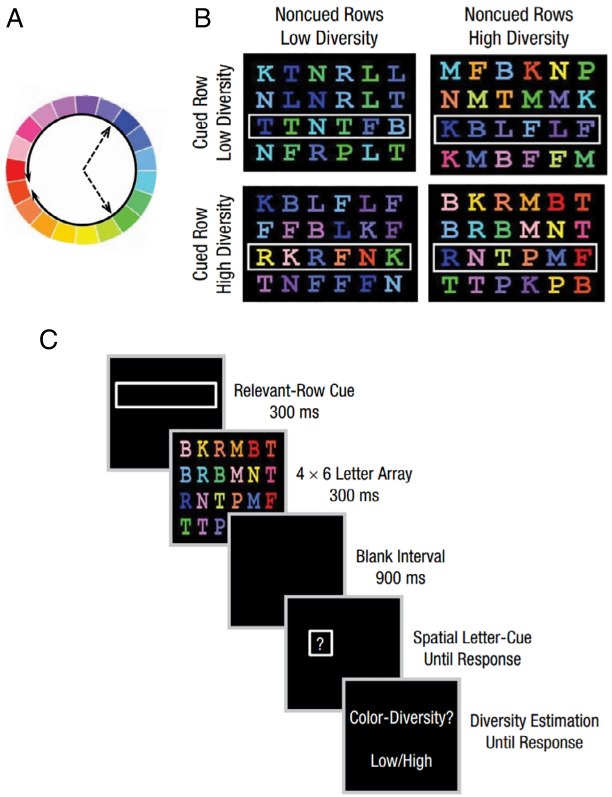

FIGURE 3 Experimental design of a modified Sperling task with color-letters, which probes the experience of colors at unattended locations. Reproduced by permission from Bronfman et al. (2014). (a) The color-wheel from which the color of the letters were randomly sampled. On high diversity trials, colors were sampled from the entire color-wheel spectrum, while on low diversity trials, the colors were sampled from a narrow range of adjacent colors (constituting 1/3 of the wheel); (b) Illustration of the four trial-types comprising two levels of color diversity (low/high) for the cued row and (independently) for the noncued rows; (c) a flow-diagram of a trial. Participants were presented with a salient cue (white rectangle), indicating the relevant row to which attention should be directed. Following the cue, a 4 × 6 color letter array appeared for 300 ms, after which a 900 ms blank interval appeared. A letter cue then indicated the letter that observers had to report. On some trials, following the letter report the observer had also indicated the color diversity of either the cued, or the noncued rows [Color figure can be viewed at wileyonlinelibrary.com]

there is no perceptual (internal) noise involved. Second, that observers are perfectly familiar with the statistics underlying the low/high CD categories (e.g., that the low-CD stimuli correspond to 1/3 of the color-range), although they received no feedback for their CD responses. Relaxing these strong assumptions by allowing some level of imprecision markedly increases the estimation of the number of unattended colors that are need in order to compute the CD. Note, however, that even if one accepts that observers experience only three colors at unattended locations at very high precision, this still validates a (thin) version of Overflow.

Second, a possible interpretation of the Bronfman et al. results is to acknowledge that observers have a generic experience of the CD at unattended locations, in terms of a summary-statistic, but to insist that this takes place in the absence of *any* awareness of individual elements (Cohen et al., 2016; Ward et al., 2016). This builds on the well-established data showing that observers can evaluate the average of a set of briefly presented elements (e.g., circles of varying magnitude; Ariely, 2001; Chong & Treisman, 2005) with being able to detect the presence of a particular element in the set.

Recently, a number of Overflow opponents (Cohen et al., 2016; Ward et al., 2016) have adopted this idea, by suggesting that summary-statistics underlie a type of generic phenomenal experience that one can have outside focal attention. In support of this position, Ward et al. (2016) reported that change blindness takes place in a CD paradigm similar to that of Bronfman et al. (2014). Participants cannot detect the shuffling of the colors in two frames separated by a brief gap.[8]

While we do not dispute that summary statistics are part of the visual experience outside attentional focus, we do contend that they are grounded on a transient experience of the constitutive elements, which is fragile and thus not encoded in working-memory and is thus unavailable for report. The difference between our view and that advanced by Overflow opponents regards the nature of the generic experience that involves summary statistics: are the summary statistics grounded on an experience of visual elements, or are they based on a computation carried out on unconsciously represented elements giving rise to a solely generic experience? Our aim below is to argue against the former interpretation and to provide support in favor of the Overflow position.

Consider first the change-blindness challenge. We wish to note here two reasons that may allow a mild version of the Overflow to resist this. First, the phenomenal experience outside focal attention is, according to Overflow, not only less detailed than the one at attended locations (thus some local details could be missed when unattended), but also very fragile (Sligte et al., 2008). The latter fact implies that once the phenomenal experience has changed, the previous visual information is erased, and therefore observers have no “backup” to use in a comparison process. Second, a mild Overflow can accept that attention is likely to be necessary for some complex visual processes, such as binding, and for the identification of detailed and complex representations such as letters, but can insist that visual attention is not necessary for experiencing visual elements such as colors and simple shapes (e.g., a rectangle). It is thus possible, that in the absence of focal attention, observers detect the presence of a variety of colors (high/low CD) but are not able to resolve their binding to specific locations, and therefore fail to detect a change that involves permutations between the colors (Ward et al., 2016). We predict, however, that observers would be able to detect that a set of new colors have been entered into the display (red/green), instead of (blue/yellow), both of which with high-CD.

Furthermore, there is one important property of CD—the summary statistic that observers experience outside focal attention—which was neglected in previous discussions of these results (Cohen et al., 2016). Unlike a first-order statistic, such as a set-average, the CD is a second-order statistic, analogous to variance (see Julesz, Gilbert, Shepp & Frisch, 1973). This is relevant to the Overflow debate, because there is a simple neural mechanism, population-averaging (Brezis, Bronfman & Usher, 2015; Georgopoulos, Schwartz & Kettner, 1986; Pouget, Dayan & Zemel, 2003), which can recover the average of a set from a highly degraded (low precision or blurred) representation of elements. According to the no-Overflow proponents, only such degraded information about visual elements is consciously available. However, because the CD is a second-order statistical property, it requires a much more precise encoding of the elements, in order to be recovered with a reasonable precision.

We demonstrate this in Figure 4, which shows an ideal-observer simulation that is based on a single parameter—the precision with which a single color element is encoded (*x*-axis, which has a cyclic range of 18 colors on a color wheel—each color corresponding to 20° on the color wheel).[9] The

[8] In Ward et al., 2016 the observers were engaged with the typical dual task (letters and CD) as in Bronfman et al. (2014), yet with one crucial addition. The spatial arrangement of the colors outside the cued row was shuffled during the trial after 100 ms from the array's onset. Importantly however, the actual color display remained unchanged (hence the CD level was unchanged); only their spatial locations changed.

[9] The assumption of a noisy variable is a simplification. In Bronfman et al. (2014; Suppl.) we proposed that color s are represented by a set of broadly tuned detectors with a ring like topography. Precision then corresponds to the sharpness of the neural profile, so that

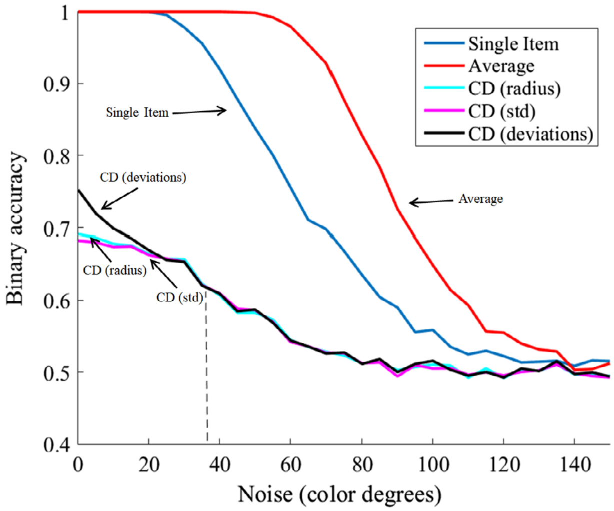

FIGURE 4 A population-code ideal-observer bound on precision. The simulation assumes a set of $n = 18$ color detectors, which are arranged around a circle (every, 20°, for a full 360°). Precision is simulated by adding a Gaussian variable to a specific color (all wrapped around the circle). The blue line shows the impact of reduced precision on a single element (binary choice). If a set of 18 elements are used instead and the color is decoded via population averaging (vector summation) the precision is enhanced to more than 95% (red line). The other three lines correspond to the impact of a single-element precision on the ideal-observer color-diversity (CD)-performance, based on three different measures for Exp 3 reported in Bronfman et al. The three measures used are: (a) sum of distances-square (black line), (b) the magnitude of the population-vector (cyan), and (c) the *SD* of the population-vector (magenta) [Color figure can be viewed at wileyonlinelibrary.com]

simulation sets a bound on the detection accuracy that can be achieved for the binary discrimination of a single color, of the average color and of the CD-performance. The binary discrimination of a single color (blue line) is defined as the probability of correctly identifying the single color's half-circle range (9/18). Here a reduction of precision corresponding to ±60° results in a performance of about 70%. The red-line shows the impact of the same reduction on single element color-precision, now on a population estimate of the color based on 18 (independently) noisy elements. The average color is computed by vector-summation on the color-ring (we assume that observers use the polar coordinate of the vector summation as their best estimate). We see that (for the same element-precision) the population color-estimate is much better than the single element one (performance exceeds .95).[10] Finally, an upper bound on the CD-performance is obtained for Experiment 3 in Bronfman et al. (2014) Three measures are used, which can distinguish between the high/low CD-conditions: (a) the sum distances square (on the ring) between all pairs of element-colors in the set; (b) the magnitude of the population-vector (cyan); (c) the *SD* of the population-vector angle (magenta). In all cases the CD score is computed, based on ideal-observer assumptions, by computing the probability of correctly identifying whether the CD score (for one of the three measures above) is above or below the median level, based on 1,000 trials, half sampled from the low and half from the high-CD conditions (Experiment 3); note that less "optimal" values of the separation criterion would only make the minimal precision estimate higher. Here we observer that in order to obtain an accuracy consistent

imprecise colors have broad/undifferentiated profiles. Based on Bayesian neural theory we showed that the sharpness of the neural profile determines in precision of the color s by a neural decoder.

[10] Note also that we do not argue that a set of degraded colors is phenomenally experienced only based on the angle of the population code vector summation. Other measures, such as the vector magnitude are likely to have an impact on the color-experience (e.g., when the visual experience of a low-magnitude population vector may correspond to "darker" or less sharp colors. However, the angle of the vector summation is still the best (ideal observer) estimate.

with the data (~66% Exp 3 in Bronfman et al., 2014) the precision needs to be *at least* as good as that of two color-detectors on the population ring (separation <40°).

To summarize, our results indicate that observers experience and discriminate a second-order statistical summary, CD, outside focal attention, whose computation requires *differentiated* (not too broad) representations of the elements. The Overflow interpretation, which we support, is that this experience is grounded on a transient (but fragile) awareness of the visual elements. Note that we do not contend that the precision with which observers experience the specific colors is the same at attended and unattended locations; we are happy with the weaker claim that they experience some of the specific unattended colors with only enough precision as to allow them to accurately report the CD of the array. Importantly, this is enough to support a mild version of Overflow, as subjects are unable to report any individual colors at the unattended rows. By contrast, the no-Overflow account of our result requires that the generic CD experience is not accompanied by any differentiated color experience of the elements, and that its computation, which requires high precision representations of the specific colors of individual letters is entirely unconscious (Phillips, 2016; Richards, 2015).

We now turn to some further (theoretical–phenomenological) considerations (section 4.1), as well as to a number of considerations from cognitive science and neuroscience (section 4.2), which, we argue, make the Overflow position more plausible, in the sense advocated by Block, of inference to the best explanation.

4 | TIPPING THE BALANCE FOR OVERFLOW

We examine here two ways to provide further support for the Overflow view, based on the CD experiment discussed above. The first involves a criticism of the notion of solely generic CD-state, and the second a criticism of the association of high-resolution color representations with unconscious perception.

4.1 | The alleged generic CD state

The notion of a representation with a solely generic phenomenology, as applied to Landman and Sperling experiments, is one in which the subject is still phenomenally aware (at least of some) of the visual properties of the unattended items, even though she is aware of them in a less determinate way. That is, there is *a way it is like for the subject* to be visually aware of these items—there is a way these items look to her—as far as the solely generic phenomenology strategy is concerned. The answer to how the specific items look to the subject can be read off of the specification of the solely generic property: they look rectangular in the one case, and they look letter-shaped in the other. The same applies to those cases best depicted by the determinate-determinable model—a particular shade of red (say, red37), so the answer goes, looks red (or, say, as falling within the range of red30 and red44). A corresponding answer for the case of solely generic CD is needed. *What does it look like to have a CD experience, without seeing (at least some of the) individual colors?*

The most supported no-Overflow proposal for the Generic state of CD appears to forgo the determinate-determinable model and the idea that the individual colors are phenomenally represented indeterminately. The suggestion is that the individual letters are phenomenally represented as different from one another with respect to their colors, though none of them is phenomenally represented as having any particular—*even if maximally indeterminate*—color. This suggestion amounts to the claim that we experience the relation of "having different colors"—a summary statistic (of difference)—without experiencing any nonrelational color-related property of the relata (i.e., the colored elements). Accordingly, the most that can be said about the looks of the relata, namely, the elements in the array, is that they are experienced as elements in a diversely colored array—they look diversely colored. On our view, in contrast, visually experiencing relations of similarity and difference vis. a vis. a certain (elementary visual) property such as color, requires visually experiencing something (nonrelational) about the relata that is relevant to that property.

We contend that the purely relational notion of a CD experience is counter-intuitive in that it radically departs from our phenomenological, introspective observations of CD experiences. We admit that, by themselves, such observations are far from conclusive, as they are based on cases in which the colors (and so CD) are attended. But, as we will attempt to show, bringing to the fore how radically different the postulated CD experiences are from familiar CD experiences, as well as exposing the theoretical commitments implied by their postulation, tell against their postulation.

There are two claims about the (unattended) CD experience that we believe both sides in the Overflow debate should accept. Hence, we take them to be constraints. First, the experience is one in which CD is *visually experienced.* It isn't that the subjects (only) have an a-modal cognitive state, such as an occurrent belief or a guess that represents that the relevant row is exhibiting CD, rather they experience it, and specifically, visually experience it, as such. Second, the experience is one in which what is visually experienced is *color* diversity. That is, subjects do not just experience that items are different from one another—the difference is not visually experienced as a "pure difference."[11] Rather, the items are visually experienced as different in a particular respect—specifically, in their *color* properties. Needless to say, the Overflow position can easily accommodate and explain both constraints. This is because, according to it, individual colors are visually experienced (even though briefly and perhaps with some imprecision), in a manner that suffices to ground the CD judgment. In contrast, we shall now argue, the two constraints are harder to accommodate in the framework of the no-Overflow position.

Note first, that in order to account for the fact that the difference in CD is *visually experienced* as a difference in *colors* (and not as a "pure difference"), the no-Overflow position is committed to postulating a basic *sui generis* visual phenomenal representation of CD. This representation is basic and *sui generis* in that its phenomenology is not a function of the visual phenomenology of color experiences.[12] The *sui generis* character of the postulated phenomenal representation follows from the view's commitment to the idea that CD is not phenomenally represented in virtue of the phenomenal representations of the individual colors (individual colors, recall, are not phenomenally represented). As such it seems to imply that the only reply to the question why is the phenomenal representation one of different *colors* is a causal, etiological answer: it is the result of (unconscious) computations on (unconscious) representation of colors. At the very least (and this suffices for our purposes), it implies that the fact that the array looks as consisting of many different colors, cannot be explained by appealing to the (if only highly indeterminate) way any individual color looks, thus suggesting that no reply to the question of "why visual *color* diversity" is available in terms that belong to the phenomenal level.

We contend that, whereas the unavailability of a reply of this sort may seem plausible in the case of primary visual elements such as the phenomenal representation of a specific color, it is more problematic with respect to the phenomenal representation of visual composites, such as color

[11] There are pop-out experiments in which the observer detects a local-difference without being able to identify the pop-out element itself (Sagi & Julesz, 1985). In such cases, however, the observer experiences a location (where) signal. Thus, even such cases do not qualify as exemplifying a "pure difference" experience.

[12] Note that the claim that the phenomenology *is* a function of the visual phenomenology of the elements leaves open the possibility that the phenomenology of the CD experience as a whole is something over and above the phenomenology of the elements—the function need not be a simple function, such as a the "sum" of the phenomenal characters of the individual items.

combinations, as in the case of the experiences that underlie the CD judgments. In other words, we grant, for present purposes, that there is no principled difficulty with the notion that, at the phenomenological level, the fact that the phenomenal representation of red is one of red is a brute unexplainable fact—a fact about which there is nothing illuminating to say in phenomenal terms. Our target is not the general claim that there must always be a story to be told at the phenomenological level, but rather the more limited claim that there is reason to expect such a story in the case of phenomenal CD representations. The reason is that the phenomenal CD representation, in contrast to the phenomenal representation of an individual color, does not seem to be "phenomenologically basic." Whereas we grant that the question why, or in virtue of what, the tomato looks red to me is bound to be left unanswered, the question why an array of items looks diversely colored to me seems different—standardly, it has an answer, namely that the array looks diversely colored because I see the individual items as having certain colors that differ from each other. So at least prima facie, the two cases (that of an individual color and that of CD) are not on a par, as far as having only an etiological explanation, and lacking a phenomenological explanation, are concerned. Admittedly, the claim that the experience of redness and the experience of CD are different in this respect is derived from the case of attended experiences. But its denial with respect to unattended CD experiences seems to point once again to how radically different from "ordinary" (i.e., attended) CD (and for that matter from imaginable) experiences the postulated purely relational CD experiences turn out to be.[13]

It seems to us that the radical difference between the attended and unattended CD experiences poses yet another explanatory challenge for the no-Overflow view. The grounding of the CD-judgments on the experience of the elements allows the Overflow position to explain what the different types of phenomenally conscious CD-states—those inside and outside the focus of attention—have in common *phenomenologically.* While *attending* to colored letters, the experience of CD is (noncontroversially) associated with the experience of the individual elements. Moreover, phenomenologically, it seems that in the attended case the phenomenal representation of CD is not *sui generis* and unstructured (as discussed above—i.e., the row looks diversely colored in virtue of seeing different individual colors). According to the Overflow position, the same kind of compositional phenomenal experience takes place outside attentional focus (but with a lower resolution; the colors are somewhat less sharp and perhaps the binding with the letters is less accurate). *A fortiori*, phenomenologically, attended and unattended phenomenal representations of CD have a lot in common—they are not totally alien. In contrast, as we saw, according to Overflow opponents, unattended CD states are quite different: they are primitive and not a function of the visual experience of the elements. This raises the question if the two different sorts of phenomenal representations of CD (attended and not attended) have enough in common to ground their apparent *phenomenological* resemblance to one another. Likewise, at the phenomenological level, is the fact that they represent the same thing—namely, they are both phenomenal representations of CD—bound to be left inexplicable? These questions are pressing, it seems to us, especially given that one of the phenomenal representations is said to be phenomenally basic and *sui generis.* Bear in mind that the wish to avoid a gap between what subjects say that they see and what they actually see is precisely what motivated the postulation of solely generic *phenomenal* representations. It remains to be seen how the no-Overflow account of (unattended phenomenal) CD states, can be developed to reply to this challenge.[14]

[13] One patient with cerebral achromatopsia—a severe impairment of color perception caused by damage in the extrastriate cortex—has been reported to be unable to identify colors, although he can detect color borders when the colors are contiguous (Heywood, Cowey & Newcombe, 1991). It can be suggested that such patients experience relational color phenomenology without experiencing individual colors, but one can also argue that they only experience the boundaries.

[14] An interesting empirical pattern also needs to be addressed. When asked to make judgments of CD at the attended row, participant's judgments are contaminated by the CD at unattended rows (Bronfman et al., 2014).

Although we have argued that the compositional Overflow account of CD experiences appears to match better with phenomenological observations than the etiological (noncompositional) account of the no-Overflow position, we believe that further experimental data is needed to provide a more decisive test between these positions. Throughout this section, we tried to strengthen the contention that v*isually experiencing* relations of similarity and difference *vis. a vis. a certain property* (in contrast, perhaps, to just experiencing a pure difference) requires visually experiencing something about the relata that is relevant to that property. This contention can be subjected to experimental verification. For example, by requesting observers to discriminate whether two elementary visual properties (patches of color (red/green) or shapes (nonsimilar letters; X/O) are the same (X/X) or not (X/0), under masking or attentional load conditions that do not allow the identification of the letters.[15] The question is whether there are stimulus to mask intervals (or load conditions) in which the participants would identify the relational property (same/different), yet would *not* identify the letters themselves? According to the no-Overflow interpretation above, this result should be achieved at some level of the stimulus to mask interval, if indeed, the summary statistic can be computed from unconscious representations of the stimuli. Conversely, according to the Overflow position we support, the performance of the participants in same/different discrimination should be accounted fully on the basis of their performance in identifying the stimuli. See Usher et al. (2018) for preliminary results.

We now turn to examine the second basic assertion of the no-Overflow position. To recap, the no-Overflow theorist accepts the existence of high precision color representations of the elements in the array, but insists that those are necessarily unconscious.

4.2 | High-precision unconscious states

The second contention of the no-Overflow account of our CD-experiment is that the high precision representations of the specific colors of unattended letters, demonstrated in our experiment, are entirely unconscious. In isolation from cognitive science literature, this may appear a reasonable position; indeed, this contention is a reiteration of the no-Overflow interpretation of the traditional Sperling experiment. We believe that one reason why many cognitive scientists are willing to entertain this position for the original Sperling experiment is that the generic conscious state suggested to underlie the introspective report for this case (array of letters)—in terms of either letter-like shapes or letter fragments (Kouider et al., 2012)—is a credible one.

An implicit appeal to Occam's razor may be a second reason for some cognitive scientists to adhere to the no-Overflow idea that considers high-resolution representations in iconic memory as unconscious in the original Sperling experiment. If we already accept (a) the presence of high-resolution representations for attended information, (b) the existence of a generic (or fragmentary) experience of shapes at unattended locations (which is consistent with data showing lower resolution outside attention, as well as the role of attention in fragment/feature binding), and (c) the existence of unconscious perceptual processes (established in many studies of subliminal cognition; reviewed in Kouider & Dehaene, 2007), why should we postulate a further component: a conscious but inaccessible high (or medium)-resolution representation (at unattended locations)?

We believe that in the case of our CD experiment there is a good reason to insist that the high resolution representations of unattended colors are transiently conscious. The key point, for which we shall now argue, is that precision is a highly diagnostic property distinguishing conscious and

[15] In order to avoid probing for binding, identification will not require distinguishing the spatial location of the stimuli (X O from O X). Increasing the stimulus set to three stimuli (e.g., X, O, T), will allow six combinations that do not distinguish between transposed items.

unconscious perception. Unlike Freudian psychology, cognitive science went through much pain before it acknowledged the existence of unconscious declarative processes (reviewed in Kouider & Dehaene, 2007). The traditional approach for demonstrating unconscious declarative cognition is to show that the participants are at chance in discriminating the relevant stimuli (Rees, Kreiman & Koch, 2002), yet there is an objective, but indirect measure of performance, such as priming, supporting the contention that the information was processed. This traditional approach was challenged on the basis of the *blindsight* phenomenon (Weiskrantz, 1986)—resulting in some accepting subjective reports as measures of conscious perception. Using the subjective report method, some studies were able to generate combined manipulations (e.g., masked stimuli with diverted attention as opposed to unmasked stimuli with full attention) that resulted in equal performance for conditions in which the stimuli are conscious or not (Rahnev et al., 2011).

While the validity of subjective reports is subject to some controversy (Schmidt, 2015), a recent integrative research article that accepts the subjective report criterion has summarized the available data concluding that "discrimination performance is typically better on *seen* than on *unseen* trials, even when the physical stimuli are physically identical" (King & Dehaene, 2014, p. 2, italics added). As the authors further state.

> …although objective discrimination can be above chance with subjectively invisible stimuli, such unconscious discrimination performance is at best mediocre. In many studies, objective discrimination performance improves dramatically when the stimuli are reported as 'seen' compared with unseen, even when sensory stimulation is identical (King & Dehaene, 2014, p. 4; see the article for additional references).

Thus, based on the recent psychophysical research, there is support for the assertion that in most cases of unconscious perception the precision of the unconscious representation is much lower than that which can be obtained under conscious perception with identical stimuli.

Further evidence based on objective measures of consciousness supporting this claim comes from a recent attentional-blink study (Asplund, Fougnie, Zughni, Martin & Marois, 2014). The attentional blink is a phenomenon that occurs when two targets are embedded in a rapid sequential visual presentation (RSVP) sequence (~100 ms/item) within a short lag of each other. Under such conditions, the detection of the first target (T1) draws attention, reducing the detectability of the second target (T2; thus the blink). Traditional attentional blink tasks, requiring report of both targets, find that at certain lags, T2 is sometimes detected (seen) and sometimes not (unseen). This raises an important question: is the distinction between seen and unseen T2s, one of graded precision? To answer this question the participants in Asplund et al. (2014) were asked to indicate the color-targets on a continuum color-wheel (guessing if they do not know), allowing thus to measure how the precision varies during the blink. The authors were able to use model fitting to contrast two hypotheses about the representation of the T2 during the blink: (a) T2 has a homogenously reduced precision, (b) T2 reduced detection is a result of averaging across two distinct states or trial types (high vs. null precision). The results clearly support the second hypothesis, namely, that unseen and seen targets are strongly distinguished in terms of their precision, even though they are physically identical. Consistent results were recently obtained with a summary-statistic task in unilateral neglect patients, who have a blind-spot in their visual field (Bisiach & Luzzatti, 1978). The results showed that unconscious information (in the neglect field) is significantly less precise than information that is consciously perceived (Pavlovskaya, Soroker, Bonneh & Hochstein, 2015).

This experimental contention, concerning the association between the precision of a neural representation and its conscious status, can be strengthened on the basis of neuroscientific considerations that involve the neural correlates of consciousness (NCC).[16] This association can be supported in two steps. The first step establishes an association between the level of activation and conscious status. We contend that at present, the best neural marker for the difference between conscious and unconscious visual perception, is the level of neural activation in visual areas. The second step establishes an association between the level of activation and the degree of precision: higher neural activation increases the representations' precision. We shall argue for each of these two steps, in turn.

Let us start with the first assertion. There are at present two dominant NCC candidates, each of which associates visual consciousness with neural activity in some specific neural circuits. The first—sometimes labeled as *local ignition*—postulates that high activation in high visual areas is the critical marker that distinguishes visual experience from unconscious perception (this activation includes local feedback loops as well as feedback to lower areas; Zeki & Bartels, 1998; Lamme, 2006; Block, 2007; Fisch et al., 2009; Noy, Bickel, Zion-Golumbic, Harel & Golan, 2015). The second—sometimes labeled as *global ignition*—postulates that the critical marker is the presence of visually induced activity in a widespread frontal–parietal network associated with the global-workspace (Baars, 1997; Dehaene et al., 2006). We contend that recent data appear to provide stronger support for the former hypothesis. Although it is beyond the scope of this paper to provide a review of this vast literature, we shall briefly allude to some crucial data regarding this debate.

The global ignition (global-workspace) NCC was supported by studies that reported activation in frontal areas that is associated with the switch of the percept during binocular rivalry[17] and is diminished during "replay" conditions that do not involve genuine rivalry (Lumer, Friston & Rees, 1998; Sterzer, Kleinschmidt & Rees, 2009). This was taken to suggest that such frontal activations trigger (and thus are causal to) the perceptual switches. However, two types of more recent findings have suggested that such frontal activations involve postperceptual processes, which are engaged by components of the tasks used to probe conscious access, but not by consciousness itself (Fisch et al., 2009; Frässle, Sommer, Jansen, Naber & Einhäuser, 2014; Tsuchiya, Wilke, Frässle & Lamme, 2015). First, in a recent review, Tsuchiya et al. (2015) highlighted an important factor—the reliance on report measures—that is likely to cause an overestimation of the NCC, as a result of the inclusion of additional processes such as monitoring and introspection. As reported by Frässle et al. (2014), when consciousness is assessed without report (e.g., using eye-tracking methods, that reliably detect the conscious percept), the differences in frontal activation at rivalry-switch between genuine and "replay" conditions disappear.

Second, the causal involvement of frontal areas has also been questioned by a transcranial magnetic stimulation (TMS) study, where stimulation of frontal areas did not reveal any disruption of conscious perception, but only impaired voluntary control of rivalry (De Graaf, De Jong, Goebel, Van Ee & Sack, 2011) and by neuropsychological data showing that a massive bilateral frontal lesion in a human patient did not abolish conscious perception, but led to deficits in cognitive, executive, visuomotor, and motivational functions (Mataró et al., 2001). In contrast, stimulation studies of high visual areas (e.g., in the face area), are typically shown to generate or distort visual (face) percepts (discussed in Koch, 2017). As Frässle et al. (2014) conclude in their review, "given the currently available evidence, activation and structural integrity of the frontal areas seems to be neither necessary nor sufficient for conscious perception" (p. 6). Based on these findings, we believe that, at present, the evidence suggests that high activity in high visual areas is the most likely neural marker of

[16] These considerations are independent of the behavioral data on the precision of conscious and unconscious percepts, discussed above, and they are relevant in the sense of "appeal to inference from the best explanation."

[17] In binocular rivalry, each eye is being presented with a different image. This procedure results in spontaneous random switches of precepts between the images (instead of perceiving their superposition).

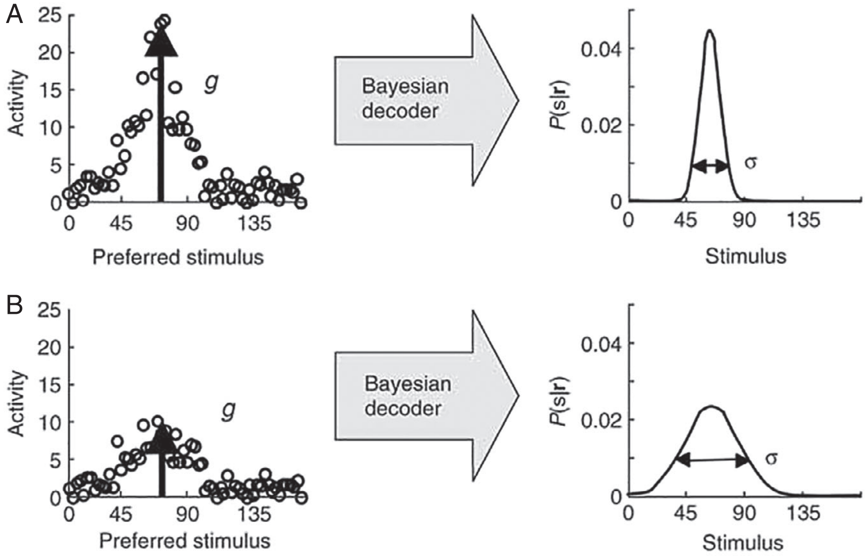

FIGURE 5 The relation between a noisy neural population activity in response to an oriented line stimulus of a specific orientation of 70° (left panels) and the posterior probability for the inferred orientation from a Bayesian model. The higher the level of activation (upper panel), the more precise is the inferred orientation of the stimulus; reproduced by permission from (Ma et al., 2006)

conscious visual experience.[18] This account is also consistent with leading theories in neuroscience, according to which consciousness relies on recurrent connections and integration of information (Lamme, 2006; Tononi, Boly, Massimini & Koch, 2016)—determinants that do not depend on access to a global-workspace.

The second step of our argument involves the association between the level of neural activity and the precision of the underlying representation. Recent research in neural coding has demonstrated that in a Bayesian framework, the level of neural activation corresponds to the degree of precision (Brezis et al., 2015; Ma, Beck, Latham & Pouget, 2006), as illustrated in Figure 5. This precision theory follows from a very simple neural principle. Neural representations are noisy, thus inherently imprecise, and their precision increases with neural activation because of an improvement in the signal to noise ratio. Hence, we can see high levels of neural activation as being necessary for precision.

If indeed, high neural activation in visual areas (which exceed an ignition threshold; e.g., Noy et al., 2015) is the most critical neural property that distinguishes conscious from unconscious perception, and if higher neural activation is a critical factor that enhances the precision of a neural representation, this provides further (neural) support for the empirical association between consciousness and precision.

As we showed in the previous section, the results of Bronfman et al. demonstrate that observers can report the CD, a second-order statistic of a set of colored letters in the absence of focal attention

[18] We acknowledge that there is a further, more philosophical, consideration against the local ignition position. This position may seem incapable of accounting for the subjective first-personal character of experiences. It may appear to leave out the defining feature of phenomenality—the "what it is like" aspect of experiences: there can be no experience that is not experienced—no feel that is not being felt; and being experienced requires an experiencer—hence, the full locution is, by necessity, "what it is like *for the subject*". The essential link between the experience and the subject, in turn, may seem to require inhabiting the global workspace, and hence frontal processing. For a review of different strategies the local ignition position can adopt in the face of this objection, see, for example, Jacobson, 2015 (section 5).

to the colors. Based on computational ideal observer considerations we argued that to make this discrimination, one has to represent the colors in a differentiated way. This interpretation receives support from experimental investigations in visual psychophysics of divided attention, which demonstrate that observers can discriminate colors (pink, orange, cyan; (Braun & Julesz, 1998) and categorize objects (Li, VanRullen, Koch & Perona, 2002; but see objection by Cohen, Cavanagh, Chun & Nakayama, 2012, and reply by Tsuchiya, Block & Koch, 2012) outside focal attention[19]; importantly the participants cannot discriminate color-conjunctions (red-green vs. green-red disks) or letters (T vs. L) under the same (divided attention) conditions (see Koch & Tsuchiya, 2007 for review). This indicates that while focal attention may indeed be necessary for binding elements, it is not necessary for conscious registration of visual elements.

While the registration of colors outside visual attention is similar in the divided attention paradigm (Braun & Julesz, 1998) and in the CD study (Bronfman et al., 2014) there is one crucial difference. In the former, as there is no WM-overload, the participants are able to report both the contents of the (central fixation) primary task and of the (peripheral) color-task. In the latter, however, the central task overloads WM, making the robust encoding of the color content into WM difficult. There is little difference between the experimental paradigms, however, with regards to the neural processes in high visual areas that mediate color discrimination, and we argue that a transient awareness of the colors outside focal attention takes place in both. In the CD task the WM-load makes the content that is experienced in the divided attention task, unavailable to report, and thus not subject to robust access.

While we do not take the results of our experiments to provide irrefutable support for the Overflow account, we believe that they provide a specific case (colors of unattended letters) that supports a mild version of Overflow, according to which there can be fleeting visual experiences of (unbound) elements outside focal attention, without robust access to working memory or report. Future research is needed to probe for stronger Overflow versions.

ACKNOWLEDGMENTS

We wish to thank Ned Block, Ian Phillips, Rafi Malach, James Stazicker and Baruch Eitam for very helpful discussions and comments on our manuscript and Noam Brezis for help with the model-simulations. All authors contributed equally to this paper. Research for this article was funded by the Israeli Science Foundation (1413/17) to MU.

REFERENCES

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Asplund, C. L., Fougnie, D., Zughni, S., Martin, J. W. & Marois, R. (2014). The attentional blink reveals the probabilistic nature of discrete conscious perception. *Psychological Science*, *25*(3), 824–831.

Baars, B. J. (1993). *A cognitive theory of consciousness*. Cambridge, MA: Cambridge University Press.

Baars, B. J. (1997). In the theatre of consciousness. Global workspace theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, *4*(4), 292–309.

Bisiach, E. & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, *14*(1), 129–133.

Block, N. (1983). Mental pictures and cognitive science. *The Philosophical Review*, *92*(4), 499–541.

Block, N. (1995). How many concepts of consciousness? *Behavioral and Brain Sciences*, *18*, 272–287.

Block, N. (2007). Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences*, *30*, 481–499.

Block, N. (2008). Consciousness and cognitive access. *Proceedings of the Aristotelian Society*, *108*, 289–317.

Block, N. (2011). Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences*, *15*, 567–575.

[19] A primary task is carried out to fully occupy focal attention

Block, N. (2012). Response to Kouider et al.: Which view is better supported by the evidence? *Trends in Cognitive Sciences*, *16*, 141–142.

Block, N. (2014). Rich conscious perception outside focal attention. *Trends in Cognitive Sciences*, *18*, 445–447.

Block, N. (2015). The puzzle of perceptual precision. In J. Windt & T. Metzinger (Eds.), *Open MIND* (pp. 1–52). Frankfurt am Main: MIND Group.

Braun, J. & Julesz, B. (1998). Withdrawing attention at little or no cost: Detection and discrimination tasks. *Attention, Perception, & Psychophysics*, *60*(1), 1–23.

Brezis, N., Bronfman, Z. Z. & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific Reports*, *5*, 10415.

Brockmole, J. R., Wang, R. F. & Irwin, D. E. (2002). Temporal integration between visual images and visual percepts. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 315–334.

Bronfman, Z. 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*(7), 1394–1403.

Chong, S. C. & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900.

Cohen, M. A., Cavanagh, P., Chun, M. M. & Nakayama, K. (2012). The attentional requirements of consciousness. *Trends in Cognitive Sciences*, *16*(8), 411–417.

Cohen, M. A. & Dennett, D. C. (2011). Consciousness cannot be separated from function. *Trends in Cognitive Sciences*, *15*, 358–364.

Cohen, M. A., Dennett, D. C. & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, *20*(5), 324–335.

De Gardelle, V., Sackur, J. & Kouider, S. (2009). Perceptual illusions in brief visual presentations. *Consciousness and Cognition*, *18*, 569–577.

De Graaf, T. A., De Jong, M. C., Goebel, R., Van Ee, R. & Sack, A. T. (2011). On the functional relevance of frontal cortex for passive and voluntarily controlled bistable vision. *Cerebral Cortex*, *21*(10), 2322–2323.

Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. New York, NY: Penguin.

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J. & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*, 204–211.

Dretske, F. (2006). Perception without awareness. In T. S. Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 147–189). Oxford: Oxford University Press.

Eagleman, D. M. & Sejnowski, T. J. (2000). Motion integration and postdiction in visual awareness. *Science*, *287*(5460), 2036–2038.

Fink, S. B. (2015). Phenomenal precision and some pitfalls: A commentary on Ned Block. In J. Windt & T. Metzinger (Eds.), *Open mind: Philosophy and the mind sciences in the 21th century* (pp. 1–14). Cambridge, MA: MIT Press.

Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., … Fried, I. (2009). Neural “ignition”: Enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron*, *64*, 562–574.

Fodor, J. A. (1975). *The language of thought*. New York, NY: Crowell, Relevant section reprinted in N. Block (Ed.). *Imagery.* Cambridge, MA: MIT Press.

Frässle, S., Sommer, J., Jansen, A., Naber, M. & Einhäuser, W. (2014). Binocular rivalry: Frontal activity relates to introspection and action but not to perception. *The Journal of Neuroscience*, *34*(5), 1738–1747.

Georgopoulos, A. P., Schwartz, A. B. & Kettner, R. E. (1986). Neuronal population coding of movement direction. *Science*, *233*, 1416–1419.

Gross, S. & Flombaum, J. (2017). Does perceptual consciousness overflow cognitive access? The challenge from probabilistic, hierarchical processes. *Mind & Language*, *32*(3), 358–391.

Heywood, C. A., Cowey, A. & Newcombe, F. (1991). Chromatic discrimination in a cortically color blind observer. *European Journal of Neuroscience*, *3*(8), 802–812.

Jacobson, H. (2015). Phenomenal consciousness, representational content and cognitive access: A missing link between two debates. *Phenomenology and the Cognitive Sciences*, *14*(4), 1021–1035.

Julesz, B., Gilbert, E. N., Shepp, L. A. & Frisch, H. L. (1973). Inability of humans to discriminate between visual textures that agree in second-order statistics—Revisited. *Perception*, *2*(4), 391–405.

King, J. R. & Dehaene, S. (2014). A model of subjective report and objective discrimination as categorical decisions in a vast representational space. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1641), 20130204.

Koch, C. (2017). The footprints of consciousness. *Scientific American Mind*, *28*(2), 52–59.

Koch, C. & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1), 16–22.

Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Psychology*, *7*(3), 341–370.

Kosslyn, S. M., Thompson, W. L. & Ganis, G. (2006). *The case for mental imagery*. Oxford: Oxford University Press.

Kouider, S. & Dehaene, S. (2007). Levels of processing during non-conscious perception: A critical review of visual masking. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *362*, 857–875.

Kouider, S., Sackur, J. & De Gardelle, V. (2012). Do we still need phenomenal consciousness? Comment on Block. *Trends in Cognitive Sciences*, *16*, 140–141.

Lamme, V. A. (2006). Towards a true neural stance on consciousness. *Trends in Cognitive Sciences*, *10*(11), 494–501.

Landman, R., Spekreijse, H. & Lamme, V. A. (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.

Leopold, D. A. & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature*, *379*, 549–553.

Lewis, K. J., Borst, G. & Kosslyn, S. M. (2011). Integrating visual mental images and visual percepts: New evidence for depictive representations. *Psychological Research*, *75*(4), 259–271.

Li, F. F., VanRullen, R., Koch, C. & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99*(14), 9596–9601.

Luck, S. J. & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Lumer, E. D., Friston, K. J. & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science*, *280*(5371), 1930–1934.

Ma, W. J., Beck, J. M., Latham, P. E. & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432–1438.

Mataró, M., Jurado, M. Á., García-Sánchez, C., Barraquer, L., Costa-Jussá, F. R. & Junqué, C. (2001). Long-term effects of bilateral frontal brain lesion: 60 years after injury with an iron bar. *Archives of Neurology*, *58*(7), 1139–1142.

Noë, A. & O'Regan, J. K. (2000). Perception, attention, and the grand illusion. *Psyche*, *6*, 6–15.

Noy, N., Bickel, S., Zion-Golumbic, E., Harel, M., Golan, T., Davidesco, I., … Malach, R. (2015). Ignition's glow: Ultra-fast spread of global cortical activity accompanying local "ignitions" in visual cortex during conscious visual perception. *Consciousness and Cognition*, *35*, 206–224.

Pavlovskaya, M., Soroker, N., Bonneh, Y. S. & Hochstein, S. (2015). Computing an average when part of the population is not perceived. *Journal of Cognitive Neuroscience*, *27*(7), 1397–1411.

Phillips, I. (2016). No watershed for overflow: Recent work on the richness of consciousness. *Philosophical Psychology*, *29*(2), 236–249.

Phillips, I. B. (2011). Perception and iconic memory: What Sperling doesn't show. *Mind & Language*, *26*, 381–411.

Pouget, A., Dayan, P. & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Review of Neuroscience*, *26*, 381–410.

Pylyshyn, Z. (2003). Return of the mental image: Are there really pictures in the brain? *Trends in Cognitive Sciences*, *7*(3), 113–118.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, *80*(1), 1–24.

Rahnev, D., Maniscalco, B., Graves, T., Huang, E., De Lange, F. P. & Lau, H. (2011). Attention induces conservative subjective biases in visual perception. *Nature Neuroscience*, *14*(12), 1513–1515.

Rees, G., Kreiman, G. & Koch, C. (2002). Neural correlates of consciousness in humans. *Nature Reviews Neuroscience*, *3*(4), 261–270.

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*(5), 368–373.

Richards, B. (2015). Advancing the overflow debate. *Journal of Consciousness Studies*, *22*(7-8), 124–144.

Sagi, D. & Julesz, B. (1985). "Where" and "what" in vision. *Science*, *228*(4704), 1217–1219.

Schmidt, T. (2015). Invisible stimuli, implicit thresholds: Why invisibility cannot be interpreted in isolation. *Advances in Cognitive Psychology*, *11*(2), 31–41.

Sligte, I. G., Scholte, H. S. & Lamme, V. A. (2008). Are there multiple visual short-term memory stores. *PLoS One*, *3*, e1699.

Sligte, I. G., Vandenbroucke, A. R., Scholte, H. S. & Lamme, V. A. (2010). Detailed sensory memory, sloppy working memory. *Frontiers in Psychology*, *1*, 1–10.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological monographs: General and Applied*, *74*, 1–29.

Stazicker, J. (2011). Attention, visual consciousness and indeterminacy. *Mind & Language*, *26*, 156–184.

Sterzer, P., Kleinschmidt, A. & Rees, G. (2009). The neural bases of multistable perception. *Trends in Cognitive Sciences*, *13*(7), 310–318.

Tononi, G., Boly, M., Massimini, M. & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience*, *17*(7), 450–461.

Treisman, A. & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, *14*(1), 107–141.

Tsuchiya, N., Block, N. & Koch, C. (2012). Top-down attention and consciousness: Comment on Cohen et al. *Trends in Cognitive Sciences*, *16*(11), 527.

Tsuchiya, N., Wilke, M., Frässle, S. & Lamme, V. A. (2015). No-report paradigms: Extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences*, *19*(12), 757–770.

Tye, M. (2006). Nonconceptual content, richness, and fineness of grain. In T. S. Gendler & J. Hawthorne (Eds.), *Perceptual experience* (pp. 504–530). Oxford: Oxford University Press.

Usher, M., Bronfman, Z. Z., Talmor, S., Jacobson, H. & Eitam, B. (2018). Consciousness without report: insights from summary statistics and inattention 'blindness'. *Philosophical Transactions of the Royal Society B*, *373*(1755), 20170354.

Ward, E. J., Bear, A. & Scholl, B. J. (2016). Can you perceive ensembles without perceiving individuals? The role of statistical perception in determining whether awareness overflows access. *Cognition*, *152*, 78–86.

Weiskrantz, L. (1986). *Blindsight: A case study and implications*. Oxford: Oxford University Press.

Zeki, S. & Bartels, A. (1998). The autonomy of the visual systems and the modularity of conscious vision. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *353*(1377), 1911–1914.

How to cite this article: Bronfman ZZ, Jacobson H, Usher M. Impoverished or rich consciousness outside attentional focus: Recent data tip the balance for *Overflow*. *Mind Lang*. 2019;34:423–444. https://doi.org/10.1111/mila.12217