The Methodological Puzzle of Phenomenal Consciousness

Phillips, Ian

DOI:
10.1098/rstb.2017.0347

The Methodological Puzzle of Phenomenal Consciousness

Ian Phillips

Birmingham University and University of Princeton

(Version of record to appear in a special issue of *Phil. Trans. R. Soc. B.* edited by M. Overgaard and P. Fazekas, DOI: 10.1098/rstb.2017.0347)

Abstract

Is phenomenal consciousness constitutively related to cognitive access? Despite being a fundamental issue for any science of consciousness, its empirical study faces a severe methodological puzzle. Recent years have seen numerous attempts to address this puzzle, either in practice, by offering evidence for a positive or negative answer, or in principle, by proposing a framework for eventual resolution. The present paper critically considers these endeavours, including partial-report, metacognitive and no-report paradigms, as well as the theoretical proposal that we can make progress by studying phenomenal consciousness as a natural kind. It is argued that the methodological puzzle remains obdurately with us and that, for now, we must adopt an attitude of humility towards the phenomenal.

Keywords: Phenomenal Consciousness, Cognitive Access, Methodological Puzzle, Phenomenal Overflow, Partial-Report Paradigms, No-Report Paradigms

The Methodological Puzzle of Phenomenal Consciousness

At the *Association for the Scientific Study of Consciousness* in 2012, Ned Block confidently wagered that disputes over whether phenomenal consciousness constitutively requires cognitive access would be settled within the decade. Since then, much innovative work has been undertaken. Yet no consensus has emerged. This reflects a deep methodological puzzle confronting consciousness science which Block himself highlights [1,2]. The study of consciousness must begin with putative cases of consciousness and unconsciousness. However, the evidence used to identify such cases (e.g. verbal report or intentional action) is equally evidence of the presence or absence of cognitive access. Thus, all our initial cases of consciousness will be presumptive cases of both consciousness and access, or of neither. Given this starting point—so the puzzle goes—how could we ever establish whether consciousness can occur without access?

The present paper offers a critical review of recent experimental and theoretical responses to the puzzle. Section one clarifies the issue at the centre of recent disputes. Section two reviews and extends earlier criticisms of partial-report studies commonly put forward as evidence of consciousness without access. Section three explains why such criticisms equally apply to studies intended to support the contrary claim that consciousness requires cognitive access. Section four challenges the contention that no-report paradigms can help resolve our quandary. Finally, section five, offers a sceptical assessment of an important theoretical framework intended to overcome the methodological puzzle due to Shea [3].

1. The Access Hypothesis

Consider a state of a subject, *S*, with content, *p*. To say that phenomenal consciousness constitutively requires cognitive access is to impose a condition on *S* being a conscious state. Current debate focuses on the following condition [2,4,5].

Access Hypothesis *S* is a conscious state only if its content *p* is "directly" available to its subject (that is: exploitable without the need for any further processing) to perform a wide-range of cognitive tasks such as reporting that *p*, or reasoning or acting on the basis of *p*.

More or less demanding access hypotheses can be formulated. More strongly, one might insist that S's content must actually be exploited for S to be conscious. More weakly, one might drop the requirement of "direct" availability. Here I focus on the Access Hypothesis as stated.

Pressing a version of the methodological puzzle, Cohen and Dennett contend that because the hypothesis of consciousness without access "cannot be empirically confirmed or falsified" [6, p. 358], it is unscientific and so "doomed" [p. 363]. However, Cohen and Dennett's considerations at best establish that the hypothesis of conscious without *function* is unscientific.[i] Despite the impression they give, such a claim falls far short of the Access Hypothesis. This is for two reasons. First, the cognitive functions mentioned in the Access Hypothesis do not exhaust all psychological functions. For example, Cohen and Dennett hold that "affective, emotional or 'limbic' reactions are … types of functions" by which the presence of consciousness could be evidenced [p. 361]. But these are not themselves cognitive functions as construed by the Access Hypothesis. Second, one can reject the Access Hypothesis on the basis that some conscious contents are only indirectly available. Yet indirect availability remains a functional characterisation (cf. [7]).

The Access Hypothesis maps onto well-established views concerning the neural and informational underpinnings of cognitive access. For example, in developing their influential global neuronal workspace model, Dehaene and Naccache distinguish three levels of accessibility: "Some information encoded in the nervous system is permanently inaccessible (set I_1). Other information is in contact with the workspace and could be consciously amplified if it was attended to (set I_2). However, at any given time, only a subset of the latter is mobilized into the workspace (set I_3)" [8, p. 30] (see also [9]). The Access Hypothesis corresponds to the claim that only mobilized information (set I_3) is conscious. Critics of the Access Hypothesis instead contend that information which is merely in contact with the workspace (set I_2) can be phenomenally consciousness [1]. Similarly, Lamme makes a critical distinction between "Stage 3" localized recurrent processing restricted to occipito-temporal areas and which "cannot directly influence motor control and other functions necessary for direct report" [10, p. 219], and "Stage 4" widespread recurrent processing involving fronto-parietal circuits which directly supports executive functions. The Access Hypothesis corresponds to the claim that consciousness requires Stage 4, global processing. Lamme thus rejects the Access Hypothesis when he argues that phenomenal consciousness is associated with Stage 3, localized recurrent processing.

Critics of the Access Hypothesis often also contend that conscious perception is rich whereas cognition is sparse. For example, Block rejects the Access Hypothesis on the grounds that the *capacity* of perceptual consciousness exceeds or "overflows" the capacity of cognitive access. However, the claim that conscious perception is rich is not, in and of itself, inconsistent with the Access Hypothesis. The apparent conflict arises from two further assumptions. First, that cognitive access is identifiable with presence in working memory. Second, that working memory has a strictly limited (say, four item) capacity.

Both assumptions can be challenged. Carruthers [4] accepts that working memory is capacity limited but denies that cognitive access equates to presence in working memory. Instead, he argues that cognitive access requires either of two forms of global broadcasting. The first form corresponds to working memory. This is capacity limited because it lacks support from bottom-up, stimulus driven activity, and so must exclusively rely on top-down attention to sustain its contents. The second form corresponds to online perception. This, Carruthers claims, allows much richer broadcast of information due to the support of bottom-up sensory activity, rendering rich perceptual consciousness consistent with the Access Hypothesis. Gross and Flombaum [11] offer an alternative way of combining rich perception with the Access Hypothesis by appeal to a conception of working memory as a continuous, flexibly-distributed and capacity-unlimited resource (see [12,13] as well as the rather different model of [14] discussed in [5]). This again affords a reconciliation of phenomenal richness with the Access Hypothesis (see also [15] discussed in [16]).

For these reasons, the main focus herein is neither overflow nor richness but the Access Hypothesis. That said, I do press the methodological puzzle by disputing studies purporting to evidence overflow since overflow is inconsistent with the Access Hypothesis. Moreover, since the theoretical contentions of Carruthers, and Gross and Flombaum remain controversial, I do not rely on either in what follows. Note though that decoupling the Access Hypothesis from overflow does not resolve the methodological puzzle. The question remains whether a state can be conscious without access, and correspondingly whether consciousness should be associated with localized as opposed to globally recurrent processing, or with being in contact with the workspace as opposed to actual mobilization into it.

2. Retrocueing Paradigms and a Recipe for Puzzlement

A substantial body of work exploiting variants on Sperling's classic partial-report paradigm [17] claims to provide evidence against the Access Hypothesis, and in favour of overflow. Phillips [18] proposes a two-step recipe for replying: (1) accept (for argument's sake) whatever interpretation is offered of the relevant data construed in purely representational or informational terms; (2) dispute the "bridging assumptions" used to move from this representational account to claims concerning consciousness.

Take Sperling's original task, widely viewed as evidence of phenomenology without access (e.g. [1,19–22] though cf. [8, p. 8] and [15] on which [16]). Following our recipe, we accept the informational import of Sperling's data, granting that they evidence a brief-lived, high-capacity "iconic memory" store selectively transferable to a stabler, low-capacity store supporting verbal report. We then deny that the full capacity of the iconic store figures in phenomenal consciousness. Instead we propose that only those contents which ultimately reach explicit or working memory do. For variations on this theme see [6,23–27].

More recently, a series of studies from Lamme's Amsterdam Group exploit a change detection task with retrocues at delays of 1–4s to argue for the existence of a fragile sensory memory store with roughly twice the capacity of working memory [28–30]. In the version of this paradigm used in Vandenbroucke et al. [31], subjects view a *memory* display of oriented rectangles for 250ms followed by a blank interval. In "iconic" and "sensory memory" conditions, this is followed by a 500ms *retrocue* highlighting the location of one of the rectangles after either 50ms (iconic condition) or 1000ms (sensory condition). After another 500ms, a *test* display is then presented in which the cued rectangle differs in orientation on half of trials. Subjects then indicate whether a change has occurred. In the "working memory" condition, the test display is shown 900ms after memory display offset, followed 100ms later, by a 500ms *postcue*. The test display then remains visible until the subject has made their

change judgment. The headline finding is that capacity (reported as number of items stored) is substantially greater in iconic and sensory as compared to working memory conditions.

This work is controversial. Critics have questioned the postulation of a distinct, high-capacity fragile memory store, either disputing the capacity claim [32] or arguing that the retrocueing effect can be understood in terms of inference and cue-driven stabilization effects *within* a single store [33,34] (see further discussion of these results in light of more recent models of working memory in [11]). However, let us set these issues aside and focus on the relation between the fragile representations postulated by the Amsterdam Group and conscious experience. The Amsterdam Group's view [29,35,10] and Block's [1,7] who follows them is that such fragile representations are conscious despite not entering (or being stabilized within) working memory. They thus contradict the Access Hypothesis.

Our recipe above provides a response on behalf of proponents of the Access Hypothesis. First, accept the existence of fragile representations as required to explain the retrocueing effect. Second, deny that all such representations correspond to elements in conscious experience. Of particular interest are the representations of items which are retrocued but were not spontaneously attended when the memory display was first shown. Cohen and Dennett [6, p. 362] claim that such representations "are stored unconsciously until the cue brings them to the focus of attention" at which point they become conscious. Phillips [24, p. 406] suggests that such representations may never reach consciousness. Instead, he suggests that, when cued, they may lead the corresponding test display rectangle to be experienced as "(un)familiar" or "(un)changed" (despite the earlier rectangle never having been consciously experienced). In the case where the rectangle has not changed orientation, experienced familiarity may reflect perceptual fluency due to prior exposure to a matching stimulus in the relevant location (cf. [36–38], and esp. [39]). Conversely, where the rectangle

has changed, a lack of fluency or perceptual "hesitancy" due to mismatch may be experienced as unfamiliarity.

Both these stories are consistent with the Access Hypothesis. Phillips' story avoids the concern that it is "implausible that unconsciously perceived stimuli can evoke conscious memories" [40, p. 223]. However, it faces its own objection, namely that subjects in a variant of the change detection task using pictures of familiar objects are significantly above chance at identifying the pre-change item from a set of four options when they successfully detect a change [30]. However, this can again be explained in term terms of fluency: subjects' previous unconscious exposure to one of the four options causes it to be experienced as more familiar than the other three items.

More recent studies from the Amsterdam Group purport to provide evidence of the association of fragile memory and consciousness (and so against the Access Hypothesis) on the grounds that metacognition is insignificantly different between fragile and working memory representations. In particular, Vandenbroucke et al. [31] (also [40]) extended the basic Amsterdam Group paradigm by asking subjects to indicate their confidence in their change detection judgement. Consistent with previous results, Vandenbroucke et al. found that memory capacity decreased from around ten items in the fragile condition to just under six in the working memory condition. Factoring in a further experiment, and exploiting a measure of metacognition, meta-*d'*-balance [41], intended to avoid the influence of varying response bias, the authors report broadly similar metacognition in both conditions. Vandenbroucke et al. conclude that "the higher capacity of fragile memory is not based on implicit, unconscious information" and thus that "sensory memory items are a meaningful part of visual experience" [31, pp. 868,870].

This interpretation assumes that if metacognitive performance concerning fragile memory is equal to that of working memory, then the information in fragile memory is conscious information. Against this, one might doubt that all working memory representations are conscious [42,43]. One might also question the association between metacognition and consciousness [44–47].

A more basic objection faces the interpretation, however. Metacognition was only measured for judgments concerning cued representations. These representations have, according to Vandenbroucke et al., been "made robust and available for report and for cognitive manipulations" [31, p. 861]. To conclude from this that there is accurate metacognition for *all* items in fragile memory requires generalizing from this cued representation. Yet strictly all that can be inferred is that "information required to support high metacognition on the entire capacity … must have been present up to the point of cue presentation" [31, p. 870]. This does not entail that subjects actually have metacognitive access to the entire capacity. Information required to detect changes in most of the rectangles must have been present up to the point of the cue. Yet plainly it does not follow from this that subjects are able to detect changes in most of the rectangles independent of the cue. Without a cue (and the attentional processing attendant on it) they mostly cannot. By the same token, we cannot assume metacognition in the absence of a cue and its attendant processing. The Access Hypothesis is thus unscathed by Vandenbroucke et al.'s data. Again we see the yawning gap between an informational story offered in explanation of certain task-performance data, and a corresponding phenomenological story. This gap precisely reflects the methodological puzzle at the heart of our discussion.

3. The Methodological Puzzle is a Two-Way Sword

This section offers two examples to illustrate that the methodological puzzle applies equally to evidence which allegedly favours the Access Hypothesis.

Sergent et al. 2013

In previous work [23,24] I suggested that Sperling's partial-report paradigm fails to provide compelling evidence against the Access Hypothesis because it is equally subject to a "postdictive" interpretation on which subjects' experiences are not determined independently of the postcue. In an elegant subsequent study, Sergent et al. [48] purport to provide clear evidence of this type of effect. Using postcues at delays of up to 400ms, they argue that "postcued attention can retrospectively trigger the conscious perception of a stimulus that would otherwise have escaped consciousness" (p. 150). However, the response recipe offered above can be used to supply an informationally equivalent but phenomenologically quite distinct interpretation of the postcueing effect found by Sergent et al. On this alternative interpretation the postcue does not trigger conscious perception but improves attention-based retention and subsequent access to already conscious experience. This interpretation (effectively the traditional interpretation of cueing in Sperling's paradigm) is consistent with theories according to which recurrent local interactions are sufficient for consciousness, and hence with access-free phenomenology.

Sergent et al. claim their "data … favor a perceptual interpretation", reasoning as follows: "Postcueing's major effect was to reduce the number of trials where participants claimed they did not see any target at all… if postcueing only improved memory of an already conscious percept, one would expect participants to shift their ratings from low, but still above 0%, visibilities toward higher visibilities, but not to change their claim of having seen the target at all" (pp. 152–3). However, there is no reason why the overflow theorist should predict shifts of the kind Sergent et al. suggest. Overflow theorists can perfectly well

hold that all-or-nothing encoding is a requirement for *reporting* visibility. This is evident if we think in terms of the response recipe provided above. Following this, the overflow theorist can simply adopt Sergent et al.'s own informational story concerning which representations are encoded for explicit report, disputing only the further claim that these are the only representations corresponding to conscious awareness.

Ward, Bear and Scholl 2016

Bronfman et al. [49], exploiting a modified Sperling paradigm using coloured letters, find that subjects can judge the colour-diversity of letters in uncued rows significantly above chance and, apparently, without cost to letter recall. They argue that this ability requires the conscious representation of the individual colours which ground the diversity judgement, and so constitutes novel evidence of overflow (see also [50]). Disputing this claim, both I [18] and Ward et al. [51] argue—in line with the recipe above—that even if the diversity judgment requires the representation of the individual colours, there is no reason (either in Bronfman's primary or supplementary data) to assume that such representations are conscious. Instead, it may simply be summary statistic representations (e.g. of "diversely coloured letters" in uncued rows) which correspond to consciousness.

Ward et al. go further, however, offering experimental evidence positively in favour of a "no overflow" interpretation of Bronfman et al.'s data. First, subjects were offered a more nuanced colour awareness scale allowing them to report: (a) no sense of colour; (b) a vague sense of colour, but not of individual letters' colours; (c) a clear sense of colour but not of individual letters' colours; and finally (d) a clear sense of individual letters' colours. Ward et al. found that most subjects claimed to perceive "color only in a general sense, without perceiving individual letters' colors" (i.e. chose options (b) or (c); p. 83). Moreover, subjects' diversity estimation was above chance just when they chose options (b)-(d) and did not

appear any more accurate when subjects chose option (d) as opposed to (b). Second, Ward et al. developed a clever change blindness paradigm in which the colours of letters in uncued rows were reshuffled on half of trials, preserving their diversity. Subjects completely failed to notice such changes despite being equally good at estimating diversity. Understandably, Ward et al. conclude that these "results are consistent with accounts of sparse visual awareness" (p. 83).

The problem with both pieces of evidence, however, is that both sides *agree* that information about individual colours is not encoded in explicit/working memory. Yet this informational claim suffices to explain why subjects will not report seeing individual colours but only colour-diversity since only the latter is encoded in explicit memory. The informational claim also suffices to explain change blindness. Change blindness (or better: difference ignorance, cf. [52]) is predicted since information about individual pre-change colours cannot be compared with information about post-change colours if it is not explicitly encoded. Change blindness is also predicted on the interpretation of change detection in retrocue paradigms mooted in Phillips [24, p. 406] where the memory display item is not encoded in explicit memory. For there too change detection depends on cue-driven attentional processing of the pre-change item(s).

It is important to recognize that this is not *ad hoc* theorizing designed to insulate phenomenal overflow from counter-evidence. Bronfman et al. themselves hypothesize that information about individual colours is not transferred "to a durable working memory store" and so "not encoded for later report" [49, p. 1395]. As a result, they ought to predict the very same data which Ward et al. find. And, indeed, Bronfman et al. make essentially this point in reply to [27]. The fact that the data cannot decide between two quite different theories here simply underscores the methodological problem at the heart of this paper—a problem which cuts both ways.

4. No-Report Paradigms

A number of authors have expressed optimism that so-called "no-report" paradigms, which attempt to investigate awareness in the absence of explicit reports, will uncover the true neural basis of consciousness, and so resolve the methodological puzzle. Tsuchiya et al., for example, emphasize no-report data in making their case that the "activation and structural integrity of the frontal areas seems to be neither necessary nor sufficient for conscious perception" [53, p. 762] (see further [54,55]). On the widely-held assumption that cognitive access is subserved by frontal areas, this amounts to the rejection of the Access Hypothesis.[ii]

In an ingenious and paradigmatic no-report paradigm, Frässle et al. [56] use two objective measures of perceptual alteration in binocular rivalry (viz. optokinetic nystagmus and pupil size) to assess the neural correlates of rivalry both with and without active report. Simplifying for argument's sake, they find that differential neural activity in frontal areas is present only in their active report condition. In their passive condition, differential activation is limited to occipital and parietal areas.[iii] Do such findings evidence that phenomenal consciousness is independent of cognitive access? I now argue that such a reaction would be precipitate (cf. [57]).

Frässle et al., in keeping with the vast majority of recent work on rivalry, are "concerned with the search for neural processes that *bring about* the spontaneous perceptual alternations that characterize multistable perception" [58, p. 81 my emphasis]. Their question is whether frontal activations cause perceptual alterations, or instead whether such alterations originate with earlier processes. This explains why the large majority of studies employ a "replay" condition in which unambiguous physical stimuli mimic perceptual alterations in the absence of rivalry processes (e.g. [59,56] and review in [58, pp. 86–8]). In analysing the relevant data, replay activation is subtracted from rivalry activation before comparing

activation in active and passive conditions. Differential activation in an area then evinces its causal role in eliciting a transition.

Our question is not this causal question, however. It is whether cognition is *constitutively* involved in consciousness, and so (granting their association) whether frontal areas form part of the constitutive basis of consciousness. The methodology of subtracting replay activity, however, means that results like Frässle et al.'s are silent on this question. For suppose frontal areas do not cause rivalry transitions, and that disambiguation occurs earlier in the perceptual hierarchy. (For evidence that this is at least sometimes the case see [60].) On this supposition, activity later in the perceptual hierarchy may well be identical in (properly matched) replay and rivalry conditions. Consequently, subtraction analysis will not reveal any differential activity. For all that, frontal activity may be a necessary condition for conscious perception.

In fact there are two possibilities to consider. First, distinguish between *core* and *total* neural correlates of a given conscious state (NCCs) [61]. A total NCC is the physical state unconditionally sufficient for being in a given conscious state. A core NCC is the part of this total realizer responsible for the state being the specific conscious state it is—crudely, its content. As just argued, results like Frässle et al.'s are quite consistent with frontal areas forming part of the *core* NCC [62]. This is because they are quite consistent with content-specific activation in frontal areas being necessary for awareness. However, even if frontal areas exhibit no content-specific frontal activity at all once activity attributable to executive upshots of awareness is factored out, frontal areas may still form part of the *total* NCC. This is because non-differential frontal activity may be a necessary condition of any non-frontal core NCC constituting a total NCC (cf. [63, p. 164] and also [8, p. 15] citing [64]). That even such an extreme finding is consistent with the Access Hypothesis highlights the limits of rivalry based paradigms in overcoming the methodological puzzle (see also [65,66]).

5. Approaching Phenomenal Consciousness as a Natural Kind

Block [1] claims that, armed with a sufficiently wide range of psychological and neuroscientific evidence, inference to the best explanation will overcome the methodological puzzle. Explicitly building on this idea, Shea presents a "systematic framework" for investigating the Access Hypothesis. The core idea of this framework is to study "phenomenal consciousness as a natural kind", thereby allowing us to "move beyond initial means of identifying instances … like verbal report … [and] find its underlying nature" [3, p. 307].

Shea's precise proposal can be summarized as follows. Our inquiry begins with defeasible evidence, E (e.g. verbal report, intentional action), for the attribution of consciousness. Based on E, we generate a large sample of putative cases of consciousness. We then examine that sample looking for distinctive neural and functional signatures or tests (T_i). Shea mentions a number of possible examples including: insensitivity to the automatic stem completion effect [67], trace conditioning [68], and gamma-band neural synchrony [69]. Finally, we exploit causal modelling techniques to search for nomological clusters amongst these signatures. A set of properties form a cluster just if "(i) they are instantiated together better than chance (given background theory); and (ii) observing subsets of the cluster supports induction to other elements of the cluster" (p. 326).

How is this procedure intended to overcome the methodological puzzle? The thought is that if we treat consciousness as a natural property then, insofar as it is not always co-instantiated with cognitive access it will have distinctive consequences which causal modelling will uncover. In this light, Shea suggests that discovering only one cluster would be "good evidence" (p. 330) in favour of the Access Hypothesis, whereas the discovery of two clusters would be "some evidence" (p. 309) against it. In this latter case, our procedure

will have arrived at a test (or battery of tests, $T_{i\text{-}j}$) which provides a better indicator of the presence of consciousness than our initial evidence E. This test will be capable of evidencing consciousness in the absence of access, thereby overcoming the methodological puzzle.

Shea's paper is ambitious and important. It deserves serious study. Here, however, I raise a series of critical issues which cast doubt on the contention that a science of consciousness which proceeds according to his framework will eventually solve the methodological puzzle.

First, Shea's proposal supposes that, at the outset of inquiry, we have evidence sufficient to provide us with samples which everyone will agree are respectively mostly conscious and mostly not conscious. It is undoubtedly true that some measures such as explicit verbal report of awareness do provide fairly uncontroversial positive evidence of consciousness. However, such superficial consensus masks the fact that even very early on in our inquiry we face profound and longstanding controversies concerning how to measure consciousness. Furthermore, it is not unreasonable to think that our initial choice of evidence will make a dramatic difference to our initial sample—a difference dramatic enough to change the number of clusters eventually found by our causal modelling. For example, consider Marcel's claim that: "There is really only one criterion for phenomenal experience. This is person's report … that they are or were conscious in one or another way…" [70, p. 131] (see also [71, p. 76] and [72, p. 1396]). Contrast this view with the "conventional" criterion for awareness adopted by many psychophysicists, namely above chance performance in a discrimination task as measured by a bias-free statistic such as *d'* [73–76]. Both "subjective" and "objective" approaches have been claimed to be traditional starting points for a science of consciousness. Moreover, quite plausibly which approach one adopts will dramatically alter the course of one's future investigation. For example, many of the tests which Shea mentions as possible differential markers of phenomenal consciousness will

count as such on a "subjective" approach but not on an "objective" approach (see, for example, [77] on insensitivity to the automatic stem completion effect, and [78] on trace conditioning). Given this, it is unclear whether all parties can even agree how to take the first step within Shea's framework.

Second, a key background assumption of Shea's approach is that cognitive access corresponds to "an information-processing mechanism … for making information directly available for use in directing a wide range of potential behaviours" (pp. 312–3). Shea takes the postulation of such a mechanism to be "plausible" (p. 314), associating it with Dehaene and Naccache's global neuronal workspace. However, he suggests: "The simplest way in which it could turn out that there is phenomenality without access … is if we discover that there is no [such] information processing property" (p. 314). We should undoubtedly be live to the possibility that there is no unified mechanism which underlies cognitive access. Dennett [79] talks of information being globally available as "fame in the brain". Since plainly societal fame is not the product of any single, unified mechanism, why not also neural fame? However, if we are rightly open to this possibility, then we must also be open to the possibility that no unified mechanism underlies phenomenal consciousness. And once this is appreciated, it becomes clear that a failure to discover a mechanism of access would not falsify the Access Hypothesis. An alternative possibility is simply that *neither* consciousness nor access have a corresponding unified, subpersonal mechanism. Furthermore, once we acknowledge the possibility that phenomenal consciousness might fail to correspond to a single, unified subpersonal mechanism, we must acknowledge that discovery of only a single kind associated with access fails to support the Access Hypothesis. For that discovery is quite consistent with access corresponding to a single, unified mechanism but not phenomenal consciousness.

Finally, and most importantly, suppose that we do in fact discover two closely connected clusters or kinds. Call these K_1 and K_2 (Figure 1(a)). In this scenario Shea suggests that we would have "reason" (p. 337) to suppose K_2 = cognitive access, and K_1 = phenomenal consciousness. This would contradict the Access Hypothesis since the merely causal connection between kinds could in principle be broken, leading to the instantiation of consciousness (K_1) in the absence of access (K_2) (Figure 1(b)).

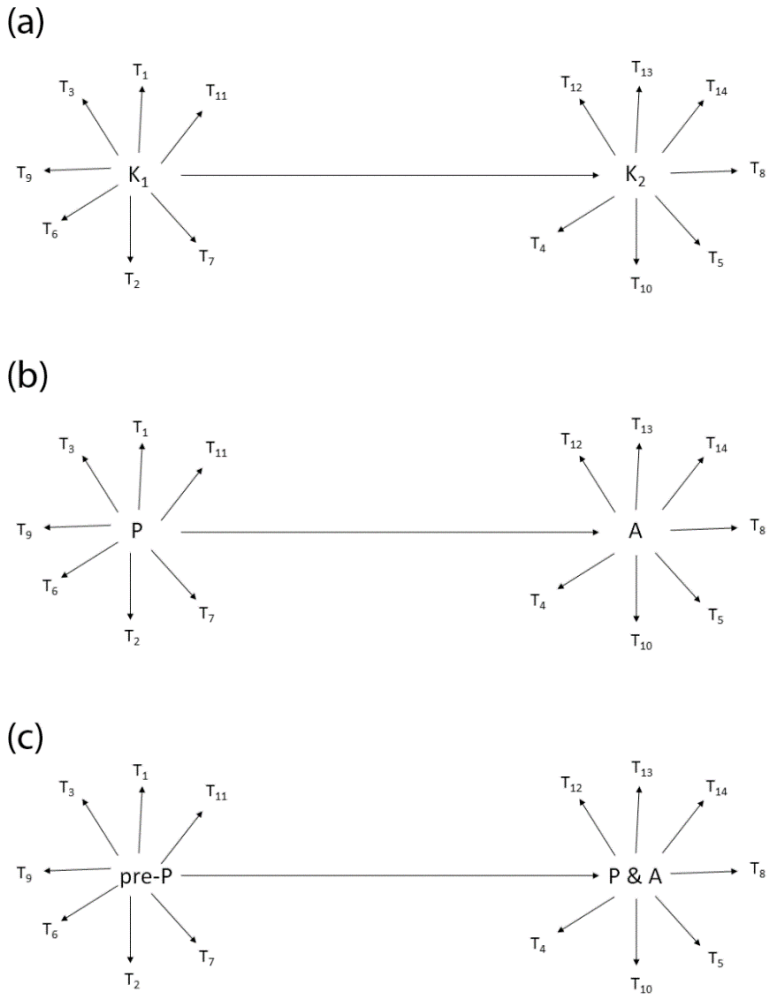

Fig. 1. (a) Hypothetical situation in which causal modelling uncovers two kinds underlying our range of putative tests for phenomenal consciousness; (b) The identification of K_1 with phenomenal consciousness, and K_2 with cognitive access; (c) The identification of K_1 with pre-consciousness and K_2 with both phenomenal consciousness and cognitive access. Based on figures from [3, p. 333].

However, recall that current workspace models postulate a distinction between information which is "in contact with the workspace and could be consciously amplified if it was attended to (set I_2)" and information which is actually "mobilized into the workspace (set I_3)". Further, recall that current disputes about the Access Hypothesis are effectively debates about whether to associate consciousness with I_2 or I_3. Thus, Block [1] claims that I_2 representations are plausibly phenomenally conscious, whereas Dehaene and colleagues [8,9] suggest that these are merely pre-conscious (supporting an illusion of rich experience) with only I_3 representations strictly being conscious. In this light, the concern naturally arises as to why we should not think that K_1 (like I_2) = pre-consciousness, and K_2 = *both* phenomenal consciousness and cognitive access (Figure 1(c)).

The point is not that being in contact with, and being mobilized into the workspace should be treated as legitimate kind properties. (Although Shea assumes that the latter is an information-based kind for the purposes of his argument, his central point is that debates about the Access Hypothesis have hitherto failed to proceed in a natural kind-based way, so it can hardly be assumed that we already know which kinds there are.) The point is rather that since defenders of the Access Hypothesis already recognize a category of pre-conscious representations, a very natural interpretation of the discovery of two clusters is open to them. On this interpretation the first kind is identified with pre-consciousness, and the second with both access and phenomenal consciousness. Given this, it is difficult to see how the discovery of two clusters would provide significant evidence against the Access Hypothesis.

Shea defends his identification of K_1 with phenomenal consciousness as follows: "Our concept [of phenomenal consciousness] refers to whatever property underpins the successful inductions in which it is deployed" (p. 335). K_1 "underpins some of those

inductions" (ibid.). Moreover, some of the clustering between our evidential tests, T_i, for K_1 and K_2 "depends on direct causal connections of some of [these tests] to [K_1]" (ibid.). It follows, Shea claims, that our concept of phenomenal consciousness refers to K_1. The problem, however, is that exactly parallel reasoning can be given for treating K_2 as the referent of our concept of phenomenal consciousness. K_2 underpins some of the successful inductions in which the concept of consciousness is deployed, and some of the clustering between evidential tests depends on direct causal connections to K_2.

In short, if we find two clusters, these will be both be directly connected to some of our putative signatures for consciousness, and jointly responsible for the normal clustering of these signatures. As a result, the proposal that our concept of phenomenal consciousness refers to whatever property underpins these successful inductions simply leaves us torn. This closely mirrors contemporary debates concerning the Access Hypothesis where theorists such as Dehaene, Block and Lamme broadly agree on the existence of two categories of representation but dispute whether the first category is phenomenal consciousness or merely pre-consciousness.

6. Conclusion

We have now reviewed both empirical and theoretical attempts to overcome the methodological puzzle facing the study of phenomenal consciousness. All have been found wanting. No argument has been given that the Access Hypothesis is beyond the reach of empirical investigation. Nonetheless, given our present data and methods, not only do we not know whether consciousness requires cognition, we do not know how to find out. Until that changes, we must adopt an attitude of humility towards the phenomenal.

Acknowledgments

Thanks to Morten Overgaard and Peter Fazekas for organizing a splendid workshop in Aarhus at which an early version of some of this material was presented, and to the other participants for lively discussion. Thanks also to Nick Shea, Steven Gross and an anonymous referee for very helpful written comments. Thanks finally to Ned Block and participants in our joint seminar at NYU for helpful feedback on various aspects of this material.

Competing Interests

I have no competing interests.

References

1. Block, N. (2007). Consciousness, accessibility and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences* 30: 481–548.

2. Block, N. (2008). Consciousness and cognitive access. *Proceedings of the Aristotelian Society* 108(3): 289–317.

3. Shea, N. (2012). Methodological Encounters with the Phenomenal Kind. *Philosophy and Phenomenological Research* 84(2): 307–344.

4. Carruthers, P. (2017). Block's Overflow Argument. *Pacific Philosophical Quarterly* 98: 65–70.

5. Gross, S. (forthcoming, this issue). Perceptual Consciousness and Cognitive Access from the Perspective of Capacity-Unlimited Working Memory. *Phil. Trans. R. Soc. B.*

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

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

8. Dehaene, S., and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* 79(1–2): 1–37.

9. Dehaene, S., Changeux, J-P., Naccache, L., Sackur, J. and Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences* 10(5): 204–11.

10. Lamme, V. A. F. (2010). How neuroscience will change our view on consciousness. *Cognitive Neuroscience*. 1(3): 204–20.

11. Gross, S., and Flombaum, J. (2017). Does Perceptual Consciousness Overflow Cognitive Access? The Challenge from Probabilistic, Hierarchical Processes. *Mind & Language* 32(3): 358–91.

12. Bays, P. M. and Husain, M. (2008). Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science* 321(5890): 851–4.

13. Ma, W. J., Husain, M., and Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience* 17(3): 347–56.

14. Oberauer, K. and Lin, H-Y. (2017). An interference model of visual working memory.. *Psychological Review* 124(1): 21–59.

15. Haun, A. M., Tononi, G., Koch, C., and Tsuchiya, N. (2017). Are we underestimating the richness of visual experience? *Neuroscience of Consciousness* 3(1): 1–4.

16. Phillips, I. B. (2018). Commentary on Haun et al. (2017): Are we underestimating the richness of visual experience? *Neuroscience of Consciousness Symposium. The Brains Blog*, forthcoming.

17. Sperling, G. (1960). The Information Available in Brief Visual Presentations. *Psychological Monographs* 74: 1–29.

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

19. Dretske, F. I. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

20. Baars, B. (1988). *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

21. Block, N. (1995). On a Confusion about a Function of Consciousness. *Behavioral and Brain Sciences* 18: 227–287.

22. Tye, M. (2006). Content, Richness, and Fineness of Grain. In T. S. Gendler and J. Hawthorne (eds), *Perceptual Experience*. Oxford: Oxford University Press, 504–530.

23. Phillips, I. B. (2011a). Attention and iconic memory. In C. Mole, D. Smithies and W. Wu (eds), *Attention: Philosophical and Psychological Essays.* Oxford: Oxford University Press, 204–227.

24. Phillips, I. B. (2011b). Perception and iconic memory. *Mind & Language* 26: 381–411.

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

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

27. 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.

28. Landman, R., Spekreijse, H., and Lamme, V. A. F. (2003). Large capacity storage of integrated objects before change blindness. *Vision Research* 43: 149–164.

29. Sligte, I. G., Scholte, H. S., and Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE* 3: e1699.

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

31. Vandenbroucke, A. R., Sligte, I. G., Barrett, A. B., Seth, A. K., Fahrenfort, J. J., and Lamme, V. A. (2014). Accurate metacognition for visual sensory memory representations. *Psychological Science* 25(4): 861–73.

32. Matsukura, M., and Hollingworth, A. (2011). Does visual short-term memory have a high-capacity stage? *Psychonomic Bulletin & Review* 18: 1098–1104.

33. Makovski, T. (2012). Are multiple visual short-term memory storages necessary to explain the retro-cue effect? *Psychonomic bulletin & review* 19(3): 470-476.

34. Griffin, I. C., and Nobre, K. (2003). Orienting attention to locations in internal representations. *Journal of Cognitive Neuroscience* 15(8): 1176–94.

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

36. Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review* 87: 252–271.

37. Jacoby, L. L., and Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General* 110(3): 306–40.

38. Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(6): 1235–1253.

39. Jacoby, L. L., and Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General* 118(2): 126–135.

40. Pinto, Y., Vandenbroucke, A. R., Otten, M., Sligte, I. G., Seth, A. K., and Lamme, V. A. F. (2017). Conscious visual memory with minimal attention. *Journal of Experimental Psychology: General* 146(2): 214–226.

41. Barrett, A. B., Dienes, Z., and Seth, A. K. (2013). Measures of metacognition on signal-detection theoretic models. *Psychological Methods* 18: 535–552.

42. Soto, D., Mäntylä, T., and Silvanto, J. (2011). Working memory without consciousness. *Current Biology* 21: R912-913.

43. King, J-R., Pescetelli, N. and Dehaene, S. (2016). Brain Mechanisms Underlying the Brief Maintenance of Seen and Unseen Sensory Information. *Neuron* 92(5): 1122–34.

44. Reder, L. M., and Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (ed.) *Implicit Memory and Metacognition*. Mahwah, N.J: L. Erlbaum, pp. 45-77.

45. Maniscalco, B., and Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness & Cognition* 21(1): 422–430.

46. Scott, R. B., Dienes, Z., Barrett, A. B., Bor, D., and Seth, A. K. (2014). Blind Insight: Metacognitive Discrimination Despite Chance Task Performance. *Psychological Science* 25(12): 2199–2208.

47. Jachs, B., Blanco, M. J., Grantham-Hill, S., and Soto, D. (2015). On the independence of visual awareness and metacognition: A signal detection theoretic analysis. *Journal of Experimental Psychology: Human Perception and Performance* 41: 269–276.

48. 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.

49. Bronfman, Z., Brezis, N., Jacobson, H., and Usher, M. (2014). We see more than we can report: 'cost free' color phenomenality outside focal attention. *Psychological Science* 25: 1394–1403.

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

51. Ward, E. J., Bear, A., and 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.

52. Dretske, F. I. (2004). Change blindness. *Philosophical Studies* 120: 1–18.

53. Tsuchiya, N., Wilke, M., Frässle, S., and Lamme. V. (2015). No-report paradigms: extracting the true neural correlates of consciousness. *Trends in Cognitive Sciences* 19(12): 757–70.

54. Storm, J. F., Boly, M., Casali, A. G., Massimini, M., Olcese, U., Pennartz, C. M. A., and Wilke, M. (2017). Consciousness Regained: Disentangling Mechanisms, Brain Systems, and Behavioral Responses. *The Journal of Neuroscience* 37(45): 10882–10893.

55. Block, N. (ms). *The Border between Seeing and Thinking*. Unpublished book manuscript.

56. Frässle, S., Sommer, J., Jansen, A., Naber, M., and Einhäuser, W. (2014). Binocular rivalry: frontal activity relates to introspection and action but not to perception. *The Journal of Neuroscience* 34(5): 1738–47.

57. Overgaard, M., and Fazekas, P. (2016). Can No-Report Paradigms Extract True Correlates of Consciousness? *Trends in Cognitive Sciences* 20(4): 241–2.

58. Brascamp, J., Sterzer, P., Blake, R., and Knapen, T. (2018). Multistable Perception and the Role of the Frontoparietal Cortex in Perceptual Inference. *Annual Review of Psychology* 69: 77–103.

59. Lumer, E. D., Friston, K. J., Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science* 280(5371): 1930–34.

60. Zou, J., He, S., Zhang, P. (2016). Binocular rivalry from invisible patterns. *PNAS* 113(30): 8408–13.

61. Shoemaker, S. (1981). Some Varieties of Functionalism. *Philosophical Topics* 12: 93–119.

62. Naber, M., and Brascamp, J. (2015). Commentary: Is the Frontal Lobe Involved in Conscious Perception? *Frontiers in Psychology* 6(1736): 1–3.

63. Jack, A. I., and Shallice, T. (2001). Introspective physicalism as an approach to the science of consciousness. *Cognition* 79: 161–196.

64. Dennett, D. C. (1991). *Consciousness Explained*. London: Penguin.

65. Blake, R., Brascamp, J., and Heeger, D. J. (2014). Can binocular rivalry reveal neural correlates of consciousness? *Phil. Trans. R. Soc. B* 369(20130211): 1–9.

66. Giles, N., Lau, H., and Odegaard, B. (2016). What type of awareness does binocular rivalry assess? *Trends in Cognitive Sciences* 20(10): 719–20.

67. Debner, J. A., and Jacoby, L. L. (1994). Unconscious perception: Attention, awareness, and control. *Journal of Experimental Psychology* 20: 304–17.

68. Clark, R. E., Manns, J. R., and Squire, L. R. (2001). Trace and delay eyeblink conditioning: contrasting phenomena of declarative and nondeclarative memory. *Psychological Science* 12(4): 304–308.

69. Crick, F., and Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in the Neurosciences* 2: 263–275.

70. Marcel, A. J. (1988). Phenomenal experience and functionalism. In: A. J. Marcel and E. Bisiach (eds) *Consciousness in Contemporary Science.* Oxford: Clarendon Press, pp. 121–58.

71. Weiskrantz, L. (1997). *Consciousness lost and found.* Oxford: Oxford University Press.

72. Naccache, L. (2006). Is She Conscious? *Science* 313(5792): 1395–1396.

73. Eriksen, C. W. (1960). Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review* 67(5): 279–300.

74. Campion, J. Latto, R., and Smith, Y. M. (1983). Is blindsight an effect of scattered light, spared cortex, and near-threshold vision? *Behavioural and Brain Sciences* 3: 423–486.

75. Campion, J., and Latto, R. (1985). What is blindsight? *Behavioral and Brain Sciences* 8(4): 755–757.

76. Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley, New York.

77. Snodgrass, M. (2002). Disambiguating conscious and unconscious influences: Do exclusion paradigms demonstrate unconscious perception? *American Journal of Psychology* 115(4): 545–579.

78. Lovibond, P. F., and Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: empirical evidence and theoretical implications. *Journal of Experimental Psychology. Animal Behavior Processes* 28(1): 3–26.

79. Dennett, D. C. (2001). Are we explaining consciousness yet? *Cognition* 79: 221–37.

80. Overgaard, M., and Grünbaum, T. (2012). Cognitive and non-cognitive conceptions of consciousness. *Trends in Cognitive Sciences* 16(3): 137.

81. Ivanova, M. V., Dragoy, O. V., Kuptsova, S. V., Yu. Akinina, S., Petrushevskii, A. G., Fedina, O. N., Turken, A., Shklovsky, V. M., and Dronkers, N. F. (2018). Neural mechanisms of two different verbal working memory tasks: A VLSM study. *Neuropsychologia* in press.

82. Fazekas, P., and Nemeth, G. (2018). Dream experiences and the neural correlates of perceptual consciousness and cognitive access. *Philosophical Transactions of the Royal Society B: Biological Sciences* this issue.

83. Wilenius-Emet, M., Revonsuo, A., and Ojanen, V. (2004). An electrophysiological correlate of human visual awareness. *Neuroscience Letters* 354: 38–41.

84. Koivisto, M., and Revonsuo, A. (2010). Event-related brain potential correlates of visual awareness. *Neuroscience and biobehavioral reviews* 34(6): 922–934.

85. Pitts, M. A., Padwal, J., Fennelly, D., Martínez, A., and Hillyard, S. A. (2014). Gamma band activity and the P3 reflect post-perceptual processes, not visual awareness. *Neuroimage* 101: 337–350.

86. Panagiotaropoulos, T. I., Deco, G., Kapoor, V., and Logothetis, N. K. (2012). Neuronal discharges and gamma oscillations explicitly reflect visual consciousness in the lateral prefrontal cortex. *Neuron* 74: 924–935.

87. Safavi, S., Kapoor, V., Logothetis, N. K., and Panagiotaropoulos, T. I. (2014). Is the frontal lobe involved in conscious perception? *Frontiers in Psychology* 5(1063): 1–2.

88. Weilnhammer, V. A., Ludwig, K., Hesselmann, G., and Sterzer, P. (2013). Frontoparietal cortex mediates perceptual transitions in bistable perception. *The Journal of Neuroscience* 33(40): 16009–15.

89. Brascamp, J. W., Brascamp, J., Blake, R., and Knapen, T. (2015). Negligible fronto-parietal BOLD activity accompanying unreportable switches in bistable perception. *Nature Neuroscience* 18(11): 1672–78.

---

[i] "At best" because it is equally unclear how the contrary hypothesis that consciousness requires access can be empirically confirmed or falsified. Thus, absent *a priori* strictures, we would seem to face an instance of underdetermination of theory by empirical data (cf. [80]) to which humility, not partisanship, would seem the rational response.

[ii] Though granted here, the assumption that frontal activity is essential for cognitive access is far from beyond question. Even on a global workspace picture, Dehaene and Naccache "see no need to postulate that any single brain area is systematically activated in all conscious states" [8, p.14], and later emphasize the contributions of neurons in inferior parietal cortex (p. 26). Moreover, recent evidence suggests that simple working memory tasks may be performed without any frontal activation (see [80], cited and discussed in [81]). Greater clarity about cognitive access and its neural basis is of critical relevance in assessing the significance of the so-called visual awareness negativity (VAN) sometimes claimed to be an index of awareness independent of global broadcasting [83–85].

[iii] Strictly, whilst Frässle et al. [56] found no significant differential activation in dorsolateral prefrontal cortex, significant activations were found in other frontal regions, including frontal eye fields and inferior frontal gyrus. This is consistent with other work suggesting differential, report-independent frontal activity in rivalry [59,86–88]. However, such activity might reasonably be argued to reflect residual executive consequences of shifts in awareness (e.g. attentional reorienting) which passive viewing fails to eliminate. This view is supported by

Brascamp et al.'s inspired "inconspicuous" rivalry paradigm in which displays of statistically and chromatically-identical quasi-randomly moving dots were used to induce unnoticeable perceptual shifts thereby minimizing executive consequences of transitions [89]. In this paradigm, no differential (switch-related) frontoparietal activity was found. Controversy remains since Brascamp et al.'s univariate voxel-wise analysis of the imaging cannot be relied on to guarantee an absence of differential activity (see also [54, p. 10884]).