



King's Research Portal

DOI:

[10.1098/rstb.2017.0350](https://doi.org/10.1098/rstb.2017.0350)

Document Version

Peer reviewed version

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Stazicker, J. (2018). Partial Report is the Wrong Paradigm. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 373(1755), [20170350]. <https://doi.org/10.1098/rstb.2017.0350>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Partial Report is the Wrong Paradigm

James Stazicker

*Department of Philosophy / Centre for Integrative Neuroscience and Neurodynamics,
University of Reading, Reading RG6 6AA, United Kingdom
ORCID: 0000-0001-6890-993X*

Keywords: partial report; consciousness; cognitive access; overflow; NCC

Is consciousness independent of the general-purpose information processes known as 'cognitive access'? The dominant methodology for supporting this independence hypothesis appeals to partial report experiments as evidence for perceptual consciousness in the absence of cognitive access. Using a standard model of evidential support, and reviewing recent elaborations of the partial report paradigm, this article argues that the paradigm has the wrong structure to support the independence hypothesis. Like reports in general, a subject's partial report is evidence that she is conscious of information only where that information is cognitively accessed. So partial report experiments could dissociate consciousness from cognitive access only if there were uncontroversial evidence for consciousness which did not imply reportability. There is no such evidence. An alternative, broadly Marrian methodology for supporting the independence hypothesis is suggested, and some challenges to it outlined. This methodology does not require experimental evidence for consciousness in the absence of cognitive access. Instead it focuses on a function of perceptual consciousness when a stimulus is cognitively accessed. If the processes best suited to implement this function exclude cognitive access, the independence hypothesis will be supported. One relevant function of consciousness may be reflected in reason-based psychological explanations of a subject's behaviour.

1. Consciousness, cognitive access and the NCC

Various theories of consciousness can be loosely grouped together as claiming that consciousness requires cognitive access [1]. 'Cognitive access' is a broad, vague term, but useful for framing the problems discussed here. It includes various processes that make information available for general purposes. For instance, cognitive access does not occur just insofar as the visual system processes information about a scene; cognitive access occurs when some of that information is made available for a range of purposes, such as verbal report, action planning, explicit reasoning and long-term mapping of the environment. So this group of theories includes, among others, theories that: consciousness consists in information's being broadcast by a system with widespread influence around the brain [2,3]; consciousness consists in a certain kind of thought about a mental representation [4]; consciousness occurs if and only if perceptual

representations of a certain kind are made available to working memory by attention [5].¹

These are all theories of what makes the difference between processing information non-consciously and being conscious of that information. Call the hypothesis that cognitive access to information is always what makes the difference between processing information non-consciously and being conscious of that information ‘the Cognitive Hypothesis’ (CH). Contrast this with theories according to which visual consciousness occurs when certain specifically visual processes occur, whether or not the visually processed information is available for general purposes [1,6,8]. According to those theories, cognitive access to information is not required for consciousness of that information.

This dispute can also be cast in neural, rather than functional, terms. According to CH, no episodes within visual cortex are sufficient for consciousness; consciousness requires projections into further mechanisms such as working memory. According to theories in the opposed group, certain episodes within visual cortex are sufficient for consciousness; for example recurrent processing—roughly, feedback loops between later and earlier areas—in visual cortex is sufficient for consciousness [6]. So this is a dispute about the neural correlates of consciousness (the NCC).

The theories in both groups are motivated by conceptual, introspective and empirical arguments. Here we are concerned only with the empirical argument against CH, though §5 and §6 explain some connections with introspective and conceptual arguments. Any empirical argument for or against CH faces a deep problem [1,9]. In controversial cases, a subject’s report of information is our only universally accepted evidence that she is conscious of that information.² Reports are ill placed to help us

¹ The claim that consciousness requires attention belongs to this group only if ‘attention’ is restricted to processes that make information available for general purposes [5,6]. Some early visual processes that meet the standard Posner criteria for attention do not satisfy this restriction [7].

² Different theorists emphasise different kinds of report [10], e.g. reporting confidence about stimuli, reporting stimulus visibility, or simply reporting stimuli themselves. Reports need not be verbal: actions such as unprompted grasping or pointing may constitute report. Other kinds of evidence for consciousness have been proposed, for example in ‘no-report paradigms’ [11], but they are controversial for reasons given in what follows.

assess CH, because whether or not consciousness requires cognitive access, reports require cognitive access: when you are in a position to report information, that information is also available for a range of other purposes. So reports are equally consistent with CH and its denial, as are failures to report.

The problem is not simply that our initial, everyday ways of recognising consciousness cannot provide evidence for or against CH. Our famously poor grasp of how physiology and information processing could explain consciousness [12,13] is a bar to moving beyond the initial ways of recognising consciousness and finding other uncontroversial ways of recognising it. Compare consciousness with information processing in this respect. Our initial way of recognising that an organism processes information typically lies in its success in navigating its environment. This does not enable us to distinguish the organism's information-processing systems from other systems, such as the bloodstream. But we can move beyond our initial way of recognising information processing to develop independent ways of identifying it. For instance, Kuffler recognised that the centre/surround structure of retinal ganglia is suitable to explain processing of information about the locations of objects' edges [14]. By contrast, because we have a poor grasp of how a physical system could explain consciousness, any non-report-based evidence for consciousness is controversial.

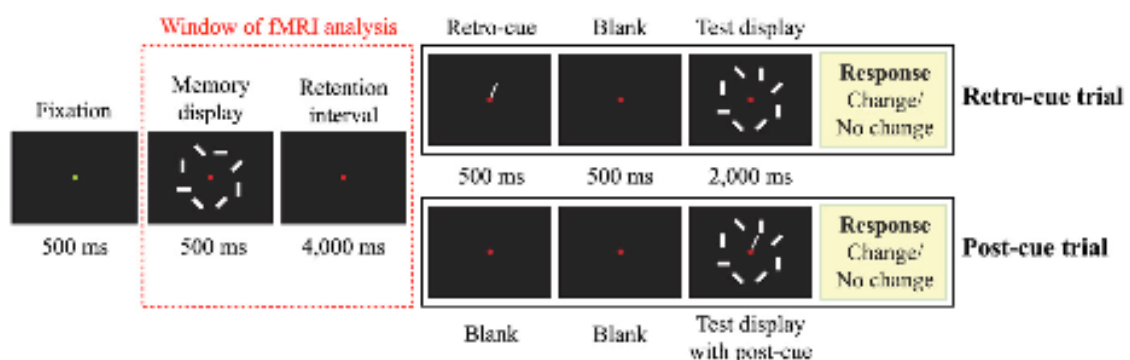
This article criticises a dominant methodology for arguing against CH. The methodology uses partial report experiments as evidence for the contrary hypothesis that perceptual consciousness 'overflows' cognitive access, in the sense that subjects are perceptually conscious of more information than is cognitively accessed [1,6,8,15-18]. Call this 'the Overflow Hypothesis' (OH). Previous work has objected that the results of partial report experiments are compatible with CH [9,19-23]. However, this objection leaves open the possibility that the evidence from partial report is a non-decisive reason for preferring OH to CH. For example, Ned Block responds that "if ... we adopt the methodology of asking which hypothesis is better supported ... we should prefer the overflow hypothesis" [16].

To assess whether partial report provides such support for OH, we can understand evidential support in the following standard way [24]. Data support a hypothesis if and only if the probability of the hypothesis, conditional on the data, is higher than the unconditional probability of the hypothesis. Conversely, data undermine a hypothesis if and only if the probability of the hypothesis, conditional on the data, is lower than the unconditional probability of the hypothesis. Finally, a hypothesis is overall well supported, given some data, to the extent to which the hypothesis is probable conditional on the data. So we can model Block’s claim as follows, where the unconditional probability of each Hypothesis is the same: conditional on the data from partial report, OH is more probable than CH. For short: $P(OH|D) > P(CH|D)$. Nothing substantive turns on this model, but it will make some distinctions clear.³

We shall see that the evidence from partial report fails to provide such support for OH, because the partial report paradigm has the wrong structure to circumvent the deep problem described above. A different methodology will be suggested.

2. Partial report

Figure 1



From Sligte *et al.* 2009 [32], by permission of Ilja Sligte and the *Journal of Neuroscience*.

³ Block’s claim is modelled in terms of absolute, rather than incremental, confirmation. That is one way (though not the only way) to capture the question of which hypothesis is best supported overall, conditional on the data from partial report.

In a series of experiments, Victor Lamme's group at the University of Amsterdam have elaborated the partial report paradigm devised by George Sperling [25]. An array of items (e.g. rectangles) is presented, followed by a blank screen, then a second array of items at the same locations (Figure 1). The task is to report whether an item in the second array differs (e.g. in orientation) from its corresponding item in the first. A line cues the relevant item. In *retro-cue* trials, this cue appears during presentation of the blank screen, up to four seconds after the first array and before the second. In *post-cue* trials, the cue appears simultaneously with the second array.

Subjects perform better in retro-cue than in post-cue trials. Performance suggests that, upon post-cuing, enough information for success in the task is available with respect to no more than four items in the first array; upon retro-cuing, enough information for success is available with respect to almost every item—exactly how many items depends on the stimuli presented and the timing of the retro-cue [26].

This is explained as follows [27]. Detailed information about almost every item in the array is processed by the visual system, and persists after offset of the array. So after offset, retro-cuing enables subjects to access detailed information about whichever item in the first array is cued. Subjects can then exploit that information to detect whether the corresponding item in the second array differs. Post-cuing, by contrast, comes too late to assist subjects in this way, because when the second array appears visual information about it overwrites visual information about the first array. As a result, in post-cue trials subjects have to rely on the limited capacity of working memory, rather than the broader capacity of specifically visual processing, to retain information about the first array.

For the sake of argument, this explanation of the results is accepted in what follows.⁴ Note that it says nothing about consciousness. So explained, the results are evidence that, after offset of the first array, detailed information about almost every item

⁴ Gross and Flombaum [28] argue that the results do not support OH, since they are explained by effects on probabilistic computations which do not involve a drop in capacity between visual processing and working memory. The argument in what follows is that the results do not support OH over CH even if they are explained by such a drop in capacity.

persists in visual system, while detailed information about ≤ 4 items reaches working memory and cognitive access.

Lamme, Block and others make the further claim that subjects are conscious of the detailed information about almost every item which persists in the visual system (OH) [1,6,15-18]. By contrast according to CH, subjects are conscious only of the information that is cognitively accessed, which includes detailed information about ≤ 4 items.

Each of these two interpretations is a straightforward prediction of the corresponding Hypothesis. OH predicts that subjects are visually conscious of information that does not reach cognitive access. CH predicts that the visual system processes information only non-consciously, where that information does not reach cognitive access. To support OH over CH, we would need to identify something that makes overflow the more probable interpretation, conditional on prior probabilities that are neutral between OH and CH.

The two interpretations are identical with respect to the structure of the information-processing hierarchy, and with respect to the level of detail processed at each stage. On both interpretations, visually processed information about almost every item is detailed enough for success in the task, while cognitively accessed information is *partial* in the following sense: this information is detailed enough for success in the task with respect to ≤ 4 items, and at best less detailed with respect to the remaining items. The interpretations differ only on the question of which processing stage correlates with consciousness.

Wayne Wu criticises the idea that subjects are conscious only of partial information on the grounds that this requires a “decay of information” not required by OH [29]. Wu argues that this commitment to a *prima facie* “pointless step” in the processing places an explanatory burden on CH. If this were correct, it would be true that $P(\text{OH}|\text{D}) > P(\text{CH}|\text{D})$. But in fact the interpretations predicted by CH and OH involve the same decay of information: that between visual processing and the lower-capacity processes of cognitive access. Similarly, Block criticises the idea that subjects

are conscious only of non-detailed information about items outside the focus of attention, on the grounds that inattention does not make early-visual responses less precise [17,30]. But CH and OH both predict that the loss of detailed information occurs at the stage of cognitive access, not in early vision.

Given this common ground between the two interpretations, any support for one interpretation over the other must concern consciousness in particular, rather than descriptions of the information processing which make no essential reference to consciousness.

One difference between the interpretations concerns how consciousness in the experiments unfolds over time. According to the overflow interpretation, retro-cuing facilitates cognitive access to detailed information about whichever item is cued, because subjects are already conscious of that detailed information before retro-cuing. That is, the overflow interpretation assumes that in retro-cue trials consciousness of the first array is cue-independent. One alternative consistent with CH is a postdictive effect such that all consciousness of the retro-cued item depends on the cue [20]. This alternative may be supported by evidence of postdictive effects on stimulus visibility even at the stimulus-cue asynchrony of 4s used in some of the Amsterdam experiments [31].

However, CH does not require that *all* consciousness of the retro-cued item depends on the cue. CH just requires that retro-cuing changes the information of which subjects are conscious in the same way as retro-cuing changes the information cognitively accessed. More specifically, retro-cuing of an item induces cognitive access to detailed information about that item; this involves a change in the information cognitively accessed by any subject who did not happen to cognitively access detailed information about that item before cuing. The overflow interpretation requires no such change in the information of which subjects are conscious. However, it requires exactly the same change in the information cognitively accessed. Since subjects report only cognitively accessed information, the two interpretations make identical predictions

about how the change affects partial reports: in the time available, detailed information about ≤ 4 items is cognitively accessed and so reported.

This problem is an instance of the more general problem facing attempts to use partial reports to support OH over CH. OH and CH agree that subjects are conscious of the partial information that is reported; the disagreement concerns unreported information. This common ground between OH and CH is not always recognised, so consider it in more detail next.

3. Consciousness explains partial reports

To cast doubt on OH, Ian Phillips notes that the Amsterdam task is two-alternative forced choice (report whether or not the cued item differs):

Two-alternative forced choice tasks notoriously over-estimate conscious experience. Information which does not correlate with consciously experienced features is often picked up by the perceptual system and encoded in implicit memory (witness priming effects well-attested across a wide range of paradigms). This information can exert an impressive influence on behavioural responses in forced-choice tasks despite absence of awareness. Thus, to establish an overflow interpretation of the results ... we need to establish that responses reflect explicit conscious comparison of the two arrays as opposed to a form of non-conscious priming. [20]

According to this priming-based alternative to overflow, subjects are not conscious of the information reflected in their partial reports.

Bronfman *et al.* [8] combined partial report with visibility ratings, to provide evidence that partial reports do reflect consciousness. They presented a 4x6 array of variously coloured letters. A pre-cue before onset indicated one row. A retro-cue indicated one letter in the same row. The task was first to report the retro-cued letter, then to report either high or low colour diversity in a certain row. High diversity

corresponds to colours spanning the entire colour wheel, low diversity to colours spanning just under one third of the wheel.

A colour-diversity task was used because “without a differentiated ... representation of the colors, it is not possible to judge diversity” [8]. This representation was dissociated from cognitive access as follows. Reports of the retro-cued letter show that subjects retained letter-identifying information about 3 letters in the pre-cued row on average. So cognitive access was focussed on the pre-cued row. Furthermore, subjects reported a row’s colour diversity equally accurately when it was the pre-cued row (69% success) and when it was not (66%). This suggests that colour-diversity reports did not depend on cognitive access to detailed information about individual items’ colours, since that would predict a greater interaction between accuracy and the focus of cognitive access. So we have evidence that only the partial information immediately involved in colour-diversity judgements reached cognitive access.

In a further experiment, subjects were given the colour-diversity task but not the letter-identification task. In addition, masking was used to reduce the array’s visibility, and subjects rated visibility. Ratings correlated with performance: ‘did not see the colors’ with 47% success; ‘partially saw the colors’ with 68%; ‘saw the colors well’ with 84%. This is evidence that subjects’ accurate partial reports about colour diversity are explained by consciousness of information about the array, rather than by non-conscious priming.

However, CH does not predict that accurate partial reports are explained by non-conscious priming. CH predicts that subjects are not conscious of information only where that information is not cognitively accessed. If the reported information about colour diversity is cognitively accessed, CH predicts that subjects are conscious of it. CH predicts only that subjects are not conscious of the underlying detailed information about individual letters’ colours.

Against CH, Bronfman *et al.* claim that subjects are conscious of the underlying detailed information. But this is not supported by the evidence that subjects exploit “a differentiated ... representation of the colors”. The issue between OH and CH is

precisely whether such representations, which are independent of cognitive access, constitute consciousness or non-conscious information processing [21]. Nor is CH undermined by the visibility ratings. According to CH, subjects are conscious only of partial information when task demands ensure that only this partial information is cognitively accessed. Subjects said they ‘partially saw the colors’ when performance (68%) was similar to performance in the experiments which loaded cognitive access with the letter-identification task (66-69%). So CH predicts the visibility ratings.

In these experiments, as in the Amsterdam experiments, reports are evidence that subjects are conscious of partial information about the array, but they are not evidence that subjects are conscious of the further information that is visually processed. Both Hypotheses predict that subjects are conscious of the partial information. The dispute concerns whether subjects are conscious of the further information. Therefore, where the data are subjects’ partial reports, $P(OH|D) = P(CH|D)$.

Some discussions and elaborations of the partial report paradigm attempt to identify non-report-based evidence that subjects are conscious of information which is not cognitively accessed. As we will see next, these attempts fail to support OH over CH, conditional on prior probabilities that are neutral between OH and CH.

4. Non-report-based evidence for consciousness?

Different forms of visual processing are at work in different partial report experiments, depending on the stimulus-cue asynchrony [26]. The system probed by retro-cues at an asynchrony of $>1s$ is dubbed *fragile Visual Short Term Memory* (fragile VSTM). In the experiment depicted in Figure 1, fMRI analysis suggested that fragile VSTM is based in area V4 of visual cortex, rather than in earlier processes at the retina, V1, V2 or V3 [32]. Block claims that this evidence “argues against highly detailed unconscious representations” of the kind required by CH, “since early visual areas would be the most obvious candidates for their locations” [16].

Block’s argument can be reconstructed as follows. Detailed unconscious representations are probable in early visual areas, but improbable in V4. Given the fMRI

evidence, CH locates detailed unconscious representations in V4, so the fMRI evidence undermines CH: $P(\text{CH} | D) < P(\text{CH})$.

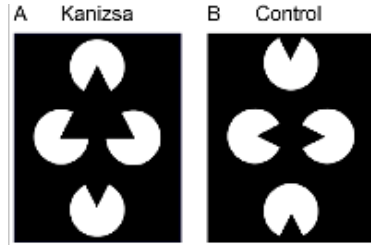
To assess this argument, consider why detailed unconscious representations may be thought improbable in V4. The information processed in early visual areas is a poor match for the information of which one is visually conscious. For example, the information of which one is visually conscious is subject to constancy effects not yet at work in V1 [33]. That makes it probable that representations in early visual areas are unconscious. But why should this, in turn, make it improbable that representations in V4 are unconscious? The only available reasoning seems to be that, if the neural correlates of visual consciousness lie somewhere in visual cortex, they probably lie in later areas such as V4. As the antecedent of this conditional makes clear, it does not undermine CH conditional on prior probabilities that are neutral between CH and OH. The case against unconscious representations in V4 is conditional on an assumption which is already inconsistent with CH—the assumption that the neural correlates of visual consciousness lie somewhere in visual cortex.

In fact, CH predicts that later visual areas such as V4 will process the detailed information that supports performance in partial report tasks. CH is the hypothesis that cognitive access to information is what makes the difference between processing information non-consciously and being conscious of that information. Aside from a possible role in criterion setting [34], the mechanisms of cognitive access are unsuitable for the further task of extracting information about distal stimuli from proximal visual stimulation. So the most probable possibility consistent with CH is that visual areas extract this information, while cognitive access just makes this information available for general purposes, which—by hypothesis—makes subjects conscious of it. Therefore, given that information processed in early visual areas is a poor match for the information of which one is conscious, CH predicts that later visual areas such as V4 will process the very information of which one is conscious.

Evidence that detailed information is processed in V4 does not, then, entail that $P(\text{OH} | D) > P(\text{CH} | D)$. Turn now to a different kind of attempt to find such support for

OH, which focuses on the character of representations in fragile VSTM, and in particular on evidence that they are governed by principles of perceptual grouping.

Figure 2



From Vandembroucke *et al.* 2012 [18], by permission of Annelinde Vandembroucke and *PLoS One*.

In a Kanizsa figure, shapes are arranged such that the visual system interprets a group of them as including an illusory contour [35]. For example, in Figure 2A a black triangle is apparent among or in front of the white shapes. In virtue of this effect, it is sometimes easier to remember the orientations of shapes which form a Kanizsa figure than to remember the orientations of shapes which do not (Figure 2B).

Vandembroucke *et al.* [18] used a Kanizsa figure as each item in the arrays of the Amsterdam paradigm. Performance was better with Kanizsa figures than with controls. Moreover, when the asynchrony between the first array and the retro-cue was $>1s$, the advantage of Kanizsa figures over controls was greater with a retro-cue than with a post-cue. This suggests that Kanizsa grouping is produced in fragile VSTM rather than working memory.

Vandembroucke *et al.* claim that their findings “suggest that the representations underlying sensory memory are phenomenological”, i.e. conscious, independently of cognitively access. To assess this, consider two different assumptions:

- (i) If fragile VSTM is a neural correlate of visual consciousness, fragile VSTM probably exhibits perceptual grouping effects.
- (ii) If fragile VSTM exhibits perceptual grouping effects, fragile VSTM is probably a neural correlate of visual consciousness.

Assumption (i) is fairly uncontroversial. Since visual consciousness exhibits perceptual grouping effects, we should expect the neural correlates of visual consciousness to exhibit them. By (i), OH predicts perceptual grouping effects in fragile VSTM. So Vandembroucke *et al.*'s findings support OH: $P(\text{OH} | \text{D}) > P(\text{OH})$. However, their findings also support CH: $P(\text{CH} | \text{D}) > P(\text{CH})$.⁵ As we saw above, the most probable possibility consistent with CH is that visual areas extract information about distal stimuli from proximal visual stimulation, while cognitive access just makes this information available for general purposes, which—by hypothesis—makes subjects conscious of it. Perceptual grouping is one aspect of how information about distal stimuli is extracted from proximal visual stimulation, so CH predicts that this will occur in visual areas. Since the findings support each Hypothesis in these ways, they do not entail that $P(\text{OH} | \text{D}) > P(\text{CH} | \text{D})$ conditional on (i).

Conditional on assumption (ii), the findings do entail that $P(\text{OH} | \text{D}) > P(\text{CH} | \text{D})$: they make it probable that fragile VSTM is a neural correlate of consciousness, and correspondingly improbable that consciousness requires cognitive access. But why accept (ii)? Vandembroucke *et al.*'s discussion suggests a motivation for doing so. They associate unconscious representation with 'fragmented' representation, and conscious representation with 'integrated' representation:

It is highly debated whether sensory memory representations are fragmented and unconscious or whether they are ... phenomenally conscious. ... Sensory memory does not merely entail simple features—such as orientation—that can also be represented unconsciously, but consists of higher-level integrated representations with a phenomenological basis. [18]⁶

⁵ Data may support both OH and CH, since OH and CH are exclusive but not exhaustive: as the history of philosophy and psychology testifies, other theories about the mechanisms of consciousness are logically possible.

⁶ This is a response to, *inter alia*, Kouider *et al.*'s proposal that fragile VSTM representations are unconscious and represent fragments of unattended items in the array [36]. That proposal is neither required nor predicted by CH (see §2). As evidence that the Kanizsa effect requires consciousness, Vandembroucke *et al.* also mention experiments which obliterated the effect through continuous flash suppression of the inducers [37]. This evidence is equivocal [21].

However, the fact that a representation is 'integrated' by perceptual grouping is not evidence that it is a conscious representation, conditional on prior probabilities that are neutral between the CH and OH. To see this, distinguish two forms of *unity* in consciousness [38]. Perceptual grouping is one, distinctively perceptual, form of unity: a group of items is represented as a unified or integrated scene; for example contours formed by the edges of several items are visible. CH is sometimes motivated by the claim that a distinct, global form of unity is a mark of consciousness: information of which one is conscious is unified, in the sense that all the information of which one is conscious at a time can be exploited together [2,3]. Since these are distinct forms of unity, the claim that global unity is a mark of consciousness incurs no commitment to the claim that perceptual unity is a mark of consciousness. In fact, to make assumption (ii) is already to assume, against CH, that what makes the difference between consciousness and non-conscious information processing is perceptual, rather than the general-purpose mechanisms of cognitive access.

Vandenbroucke *et al.* qualify their interpretation in a way that sets out the methodology of Lamme's group:

Obviously, to warrant the conclusion that sensory memory holds items in a fully perceptual or even conscious status would require more evidence. Many dimensions of perceptual quality would have to be compared between sensory memory and unequivocally conscious representations, either psychophysically or using neuroimaging techniques. Only when sensory and attended or accessed representations are similar in many or all perceptual dimensions the conclusion is warranted that sensory memory is a remnant of conscious vision. [18]

The methodology is to identify many "dimensions of perceptual quality" which, like perceptual grouping, characterise both fragile-VSTM representations and clear cases of consciousness. This methodology will not find evidence such that $P(OH | D) > P(CH | D)$,

conditional on prior probabilities that are neutral between OH and CH. Again, distinguish two assumptions:

- (1) If fragile VSTM is a neural correlate of visual consciousness, fragile VSTM probably exhibits perceptual qualities $Q_1 \dots Q_n$.
- (2) If fragile VSTM exhibits perceptual qualities $Q_1 \dots Q_n$, fragile VSTM is probably a neural correlate of visual consciousness.

Assumption (1) is fairly uncontroversial, and by (1) OH predicts perceptual qualities $Q_1 \dots Q_n$ in fragile VSTM. So the discovery of such qualities in fragile VSTM would support OH. But the discovery of such qualities in fragile VSTM would also support CH. For CH predicts that visual areas will extract information about distal stimuli from proximal visual stimulation in ways that exhibit the distinctively perceptual aspects of perceptual consciousness, while cognitive access just makes the results available for general purposes and so aspects of consciousness.

Conditional on assumption (2), the discovery of perceptual qualities $Q_1 \dots Q_n$ in fragile VSTM would entail that $P(\text{OH} \mid D) > P(\text{CH} \mid D)$. But to assume (2) is already to assume, against CH, that one mechanism probably plays two distinct explanatory roles: on the one hand, explaining the distinctively perceptual aspects of perceptual consciousness; on the other hand, making the difference between non-consciously processing information and being conscious of that information. CH is precisely the hypothesis that distinct mechanisms play these distinct roles.

As we saw in §3, where the data are partial reports, $P(\text{OH} \mid D) = P(\text{CH} \mid D)$. We have now seen two arguments—from fMRI evidence and perceptual quality—which seek non-report-based evidence that subjects are conscious of information which is not cognitively accessed. The arguments fail because there is no such evidence, conditional on prior probabilities that are neutral between OH and CH. The evidence remains such that $P(\text{OH} \mid D) = P(\text{CH} \mid D)$. Turn now to a different attempt to find evidence for consciousness beyond subjects' partial reports. This approach starts with the partial

reports themselves, and exploits an assumption about the specificity of consciousness to infer from them that subjects are conscious of more than they report or cognitively access.

5. The specificity of consciousness

Subjects' formally measured partial reports concern only the ≤ 4 items about which cognitively accessed information is specific enough for success in the task. Block adds a further, informal element: "subjects said that they could see all or almost all" the items in the array [1,2].

Some defenders of CH cast such reports as the product of introspective illusions [3,36]. If CH required introspective illusions, this might be an explanatory burden such that $P(OH|D) > P(CH|D)$. But CH requires no introspective illusions, even if Block's account of the reports is correct. Accept, for the sake of argument, that where the task is to report whether there is a difference in the orientation of a rectangle (Figure 1), subjects accurately report seeing every rectangle.

Block takes the reports to show that, for each rectangle, the subject is conscious that it is a non-square rectangle [16]. Let us agree that a subject reports seeing a rectangle only if she is conscious that it is a non-square rectangle. Since this information is reported, it is cognitively accessed. So the information cognitively accessed is as follows: for each rectangle, information that it is a non-square rectangle; for ≤ 4 rectangles, enough specific information about orientation for success in the task; for the remaining rectangles, less specific information about orientation.⁷

CH requires that this is also the information of which subjects are conscious. Block objects that "generic conscious representations of non-square rectangles that do

⁷ By Block's own lights, then, his suggestion that cognitive access "has a capacity of about 4 items" [1] must refer only to the capacity to process *detailed* information about an item. This is consistent with recent models of working memory which do not give its capacity simply in terms of a number of items [39].

not specify between horizontal and vertical orientations is difficult to accept” [16].⁸ That is, he assumes support for the following principle:

Specificity If one is visually conscious that an item is a non-square rectangle, one is also visually conscious of more specific information about its orientation.

If *Specificity* is to undermine CH, *Specificity* must include cases in which the more specific information about orientation is not cognitively accessed. So understood, *Specificity* entails OH and contradicts CH. Therefore we have already seen the following, where prior probabilities are neutral between OH and CH: any support for *Specificity* must concern consciousness in particular, rather than descriptions of information processing which make no essential reference to consciousness (§2); reports about stimuli do not support *Specificity*, since they reflect only cognitively accessed information about stimuli (§3), yet any support for *Specificity* must be report-based (§4). So any support for *Specificity* must lie not in the results of partial report experiments, but in a particular kind of introspective report: a report which provides evidence that one is conscious of specific information about a stimulus, but which does not depend on cognitive access to that specific information.

On one influential model of introspection, such an introspective report is impossible, because you introspect perceptual states by cognitively accessing the information about stimuli which those states make available [40]. Support for *Specificity* requires a different model, such that you can introspect that you are conscious of specific information about a rectangle’s orientation without cognitively accessing that information. For example, one might hold that introspective reports reflect cognitive access to information about properties of conscious perception which in turn *carry*

⁸ Block’s 2007 characterisation of what CH requires was more restrictive: “a phenomenal presentation *that there is a circle of rectangles*” [1]. Such a presentation, which does not involve seeing each item, is not required by CH. See Stazicker 2011 [9] for criticism of this restriction, and of Block’s related claim that CH requires a dramatic change from generic to specific consciousness which subjects could be expected to report.

information about the environment. On this model, introspecting a perceptual state is analogous to seeing that a tree has rings which carry specific information about its age; you can do this without seeing how many rings there are or what the specific age is.

For present purposes, we need not settle whether such an introspective report is possible, nor whether it could provide reliable evidence.⁹ Suppose there is introspective support for *Specificity*. If so, $P(OH|D) > P(CH|D)$ where the data are purely introspective. The partial report experiments are irrelevant. As we have seen, the experimental results do not support *Specificity*. Nor do the experiments add complementary evidence for OH and against CH, such that $P(OH|D) > P(CH|D)$ where the data are the experimental results together with the introspective evidence. For as we have seen, without the introspective evidence, $P(OH|D) = P(CH|D)$. So with or without the introspective evidence, the experiments do not provide evidence such that $P(OH|D) > P(CH|D)$.

6. An alternative methodology?

Partial report is a fruitful paradigm. Conditional on OH, the experiments isolate a neural correlate of consciousness in fragile VSTM, and provide important evidence about its character and function. However, the paradigm has the wrong structure to establish that $P(OH|D) > P(CH|D)$. Conditional on prior probabilities neutral between OH and CH, the experiments provide evidence only that subjects are conscious of the partial information which is cognitively accessed.

This is one instance of a more general problem with seeking experimental evidence for consciousness in the absence of cognitive access: reports are our only uncontroversial evidence of consciousness, and reported information is cognitively accessed. So on the face of it we need a methodology which does not require such evidence, if CH is to be critically assessed on empirical grounds.

⁹ See Spener 2007 [41] for doubts about reliability.

Following David Marr [42], a standard methodology for investigating a psychological system is to identify a function of the system, an algorithm that could compute the function, and physiological processes suitable to implement the algorithm. This does not require evidence that the function is performed in the absence of its downstream effects. For example, it was possible for Kuffler to identify the function of edge detection, and argue that ganglia are suitable to implement an algorithm that computes this function, without experimental evidence of edge detection in the absence of downstream effects such as object perception [14]. A similar approach to critically assessing CH would not seek evidence for consciousness in the absence of cognitive access; it would identify a function of consciousness, and assess whether the mechanisms most suitable to implement that function include those of cognitive access.

One set of obstacles to this approach lies in identifying a suitable function of consciousness. On the one hand, consciousness is sometimes said to have no function [43]. On the other hand, where consciousness is treated as having a function, CH is stipulated from the outset. For example some defenders of CH assume, in effect, that the function of perceptual consciousness takes perceptual information as input and delivers reportability as output [3]. This stipulates CH because by assumption, the function of perceptual consciousness includes a function of cognitive access, namely the kind of general-purpose availability of information required for report [1]. If CH is to be critically assessed, we need a more fine-grained account of the function of perceptual consciousness which neither assumes nor denies CH, but which is consistent with the starting-point that evidence for consciousness is report-based.

Articulating and defending such an account would involve tackling a range of broadly-speaking conceptual questions about consciousness, thought and action—a substantial project beyond our scope here. But to illustrate one example of a possible direction this could take, consider a study of consciousness in which it is initially unclear whether a response is a report.

In Adrian Owen and colleagues' famous work with outwardly unresponsive patients who meet the criteria for vegetative state, Monti *et al.* [44] asked patients to

imagine playing tennis, then to imagine walking around their homes. Several responded with neural activity that is correlated in healthy subjects with consciously imagining these actions (in Sensory Motor Area and Parahippocampal Place Area respectively). Perhaps this could be cast as a non-conscious priming effect, but a further experiment seems to confirm that, in some patients, responses of this kind are reports and reflect consciousness. Patients who responded as above were instructed to imagine playing tennis in order to answer 'yes', and to imagine walking around the house in order to answer 'no', in response to questions such as 'Do you have any brothers?' One patient's responses corresponded to correct answers.

Why is it so compelling that this patient was conscious? Monti *et al.* suggest a natural answer, characterising her responses as "willful" communication. On the face of it at least, the fact that the responses corresponded to correct answers can be explained only if they were intentional actions, reflecting a decision made on the basis of the instructions and questions. That is evidence that the patient was conscious, insofar as decisions of this kind require consciousness of the basis for action.

The suggestion that decisions of this kind require consciousness can be fleshed out using some ideas from the philosophical literature. Intentional action is often understood as action for reasons [45]. For instance the instructions and the question are among the patient's reasons for imagining playing tennis. A growing body of work explores the idea that one has a reason or rational basis for certain thoughts (perceptual demonstratives, such as those expressed by saying 'that person' or 'that instruction') only if one consciously perceives the referent of the thought [46]. Accordingly, we might express the function of perceptual consciousness appealed to here as follows. It takes perceptual information as input, and delivers as output a kind of sensitivity to reasons for thought and action which can inform a decision to act intentionally.

From this perspective, reports are evidence of consciousness because they are a way of acting for the reasons consciousness makes available. But the function of perceptual consciousness is not just a function from perceptual information to reportability, or to anything that obviously requires cognitive access. By itself, the

function identified leaves it open whether cognitive access is required for the kind of sensitivity to reasons which can inform a decision to act intentionally, or just for the decision itself. So this is an example of a function of consciousness which, if it could be defended, might allow us to critically assess CH.

A second set of obstacles to this approach lies in identifying suitable algorithms and physiological processes. To illustrate using the same example, it is unclear how the notion of an algorithm should be applied to a function which takes perceptual information as input, and delivers as output a different kind of sensitivity to that same information. Partly as a result, we are a long way from having evidence about whether the mechanisms of cognitive access, or instead earlier processes which provide input to cognitive access, such as those within visual cortex, are suitable to implement this function.

However, it is worth noting that the problem here is not the famously hard one of explaining consciousness itself in physiological terms [13]. To identify a *function* of consciousness and physiological processes suitable to implement an algorithm for computing that function, we need not overcome our poor grasp of how physiological processes could explain the character of consciousness itself. In this respect, a broadly Marrian approach has a basic advantage over attempts to identify neural correlates of consciousness independently of its function—for example, over attempts to identify consciousness in the absence of cognitive access. Nonetheless, as these brief remarks illustrate, such an approach to the empirical critical assessment of CH faces serious obstacles, and would require a challenging engagement with both philosophical and empirical problems. Perhaps that is what we should expect in an attempt to identify the mechanisms of consciousness.¹⁰

¹⁰ Work on this material was presented at meetings of the Association for the Scientific Study of Consciousness and the European Society for Philosophy and Psychology, the Oxford Mind Work in Progress Seminar, and a British Academy Newton Fund workshop at the Universidad Nacional Autónoma de México. For helpful discussions, thanks to Ned Block, Michael Caie, Tony Cheng, Will Davies, Baruch Eitam, Stephen Fleming, Anil Gomes, Matthew Parrott, Ian Phillips, Philip Stratton-Lake, Miguel Ángel Sebastián, Ilja Sligte and Marius Usher. Thanks to two anonymous referees for helpful comments.

References

1. Block, N. 2007 Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral & Brain Sciences* 30, 481-548.
2. Baars, B. 1988 *A Cognitive Theory of Consciousness*. New York: Cambridge University Press.
3. 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(5), 204-11.
4. Rosenthal, D. 2005 *Consciousness and Mind*. Oxford: Clarendon Press.
5. Prinz, J. 2012 *The Conscious Brain: How attention engenders experience*. New York: OUP.
6. Lamme, V. 2006 Towards a true neural stance on consciousness. *Trends in Cognitive Sciences* 10(11), 494-501.
7. Posner, M. 1980 Orienting of attention. *Quarterly Journal of Experimental Psychology* 32(1), 3-25.
8. 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(7), 1394-1403.
9. Stazicker, J. 2011 Attention, visual consciousness and indeterminacy. *Mind & Language* 26(2), 156-184.
10. Irvine, E. 2013 Measures of consciousness. *Philosophy Compass* 8(3): 285-297.
11. 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), 757-770.
12. Nagel, T. 1974 What is it like to be a bat? *Philosophical Review* 83, 435-456.
13. Levine, J. 1983 Materialism and qualia: The Explanatory Gap. *Pacific Philosophical Quarterly* 64, 354-361.
14. Kuffler, S. 1952 Neurons in the retina: organization, inhibition and excitation problems. *Cold Spring Harbor Symposia on Quantitative Biology* 17, 281-292.
15. Tye, M. 2006 Content, richness, and fineness of grain. In Gendler, T and Hawthorne, J (eds.), *Perceptual Experience*. Oxford: OUP.
16. Block, N. 2011 Perceptual consciousness overflows cognitive access. *Trends in Cognitive Sciences* 15(12), 567-575.
17. Block, N. 2014 Rich conscious perception outside focal attention. *Trends in Cognitive Sciences* 18(9), 445-447.
18. Vandenbroucke, A, Sligte, I, Fahrenfort, J, Ambroziak, K and Lamme, V. 2012 Non-attended representations are perceptual rather than unconscious in nature. *PLoS One* 7(11), e50042.
19. Sergent, C. and Rees, G. 2007 Conscious access overflows overt report. *Behavioral & Brain Sciences* 30, 524.
20. Phillips, I. 2011 Perception and iconic memory: what Sperling doesn't show. *Mind & Language* 26(2), 381-411.
21. Phillips, I. 2016 No watershed for overflow: Recent work on the richness of consciousness. *Philosophical Psychology* 29(2), 236-249.
22. Brown, R. 2011 The myth of phenomenological overflow. *Consciousness & Cognition* 21, 599-604.
23. Cohen, M. and Dennett, D. 2011 Consciousness cannot be separated from function. *Trends in Cognitive Sciences* 15(8), 358-364.
24. Háyeck, A and Joyce, J. 2008 Confirmation. In Psillos, S and Curd, M (eds.), *Routledge Companion to Philosophy of Science*. New York: Routledge.
25. Sperling, G. 1960 The information available in brief visual presentations. *Psychological Monographs: General & Applied* 74(11, Whole No. 498), 1-29.
26. Sligte, I, Scholte, H and Lamme, V. 2008 Are there multiple visual short-term memory stores? *PLoS One* 3(2), e1699.
27. Landman, R, Spekreijse, H and Lamme, V. 2003 Large capacity storage of integrated objects before change blindness. *Vision Research* 43(2), 149-64.
28. Gross, S and Flombaum, J. 2017 Does perceptual consciousness overflow cognitive access? The challenge from probabilistic, hierarchical processes. *Mind & Language* 32(3), 358-391.
29. Wu, W. 2014 *Attention*. New York: Routledge.
30. Ling, S, Liu, T and Carrasco, M. 2009 How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research* 49, 1194-1204.
31. Sergent, C, Wyart, V, Babo-Rebelo, M, Cohen, L, Naccache, L and Tallon-Baudry, C. 2013 Cueing attention after the stimulus is gone can retrospectively trigger conscious perception. *Current Biology* 23(2): 150-155.
32. Sligte, I, Scholte, H and Lamme, V. 2009 V4 activity predicts the strength of visual short-

- term memory representations. *Journal of Neuroscience* 29(23), 7432-7438.
33. Crick, F and Koch, C. 1998 Consciousness and neuroscience. *Cerebral Cortex* 8, 97–107.
 34. Lau, H and Rosenthal, D. 2011 Empirical support for higher-order theories of conscious awareness. *Trends in Cognitive Sciences* 15(8), 365-73.
 35. Kanizsa, G. 1976 Subjective contours. *Scientific American* 234, 48–52.
 36. Kouider, S, De Gardelle, V, Sackur, J and Dupoux, E. 2010 How rich is consciousness? The partial awareness hypothesis. *Trends in Cognitive Science* 14, 301–307.
 37. Harris, J, Schwarzkopf, D, Song, C, Bahrami, B and Rees, G. 2011 Contextual illusions reveal the limit of unconscious visual processing. *Psychological Science* 22, 399–405.
 38. Bayne, T. 2010 *The Unity of Consciousness*. Oxford: OUP.
 39. Ma, W, Husain, M and Bays, P. 2014 Changing concepts of working memory. *Nature Neuroscience* 17, 347–56.
 40. Evans, G. 1982 *The Varieties of Reference*. Oxford: Clarendon Press.
 41. Spener, M. 2007 Expecting phenomenology. *Behavioral & Brain Sciences* 30, 526-527.
 42. Marr, D. 1982 *Vision*. San Francisco: W.H. Freeman.
 43. Rosenthal, D. 2008 Consciousness and its function. *Neuropsychologia* 46: 829-840.
 44. Monti, M, Vanhaudenhuyse, A, Coleman, M, Boly, M, Pickard, J, Tshibanda, L, Owen, A and Laureys, S. 2010 Willful modulation of brain activity in disorders of consciousness. *New England Journal of Medicine* 362(7), 579-589.
 45. Anscombe, GEM. 1957 *Intention*. Oxford: Basil Blackwell.
 46. Dickie, I. 2015 Perception and demonstratives. In Matthen, M., *Oxford Handbook of Philosophy of Perception*. Oxford: OUP.