Comment on BBS target article by Firestone & Scholl

**Word count**

Abstract (max. 60): 59  
Main text (max. 1k): 1000  
References: 844  
Entire text: 1992
Attention and multisensory modulation argue against total encapsulation

Benjamin de Haas, Dietrich Samuel Schwarzkopf, Geraint Rees

Institute of Cognitive Neuroscience, 17 Queen Square, WC1N 3AR, Wellcome Trust Centre for Neuroimaging, 12 Queen Square, WC1N 3BG, Experimental Psychology, 26 Bedford Way, WC1H 0AP

University College London, UK

Telephone: +44-20-7679-5349

E-mail:

benjamin.haas.09@ucl.ac.uk
s.schwarzkopf@ucl.ac.uk
g.rees@ucl.ac.uk

Web:

https://www.ucl.ac.uk/pals/research/experimental-psychology/person/benjamin-de-haas/
https://sampendu.wordpress.com/sam-schwarzkopf/
http://www.fil.ion.ucl.ac.uk/~grees/

Abstract

Firestone and Scholl postulate that vision proceeds without any direct interference from cognition. We argue this view is extreme and not in line with the available evidence. Specifically, we discuss two well-established counter-examples: attention directly affects core aspects of visual processing, and multisensory modulations of vision originate on multiple levels, some of which are unlikely to fall ‘within perception’.

Main Text

Firestone and Scholl argue there is no good evidence for cognitive penetration of perception, specifically vision. Instead, they propose, visual processing is informationally encapsulated. Importantly, their version of encapsulation goes beyond Fodor’s original proposal ‘that at least some of the background information at the subject’s disposal is inaccessible to at least some of his perceptual mechanisms’ (Fodor, 1983). Their hypothesis is much more ambitious: ‘perception proceeds without any direct, unmediated influence from cognition’. We will refer to this view as total encapsulation.
One possible counterexample to total encapsulation is *multisensory* modulation. For example, sounds in rapid succession can induce the illusory re-appearance of visual flashes (Shams, Kamitani, & Shimojo, 2000). Such re-appearances increase objective sensitivity for visual features of the flash (Berger, Martelli, & Pelli, 2003) and are linked to individual structure and function of primary visual cortex (de Haas, Kanai, Jalkanen, & Rees, 2012; Watkins, Shams, Tanaka, Haynes, & Rees, 2006). Waving one’s hand in front of the eyes can induce visual sensations and enable smooth pursuit eye movements, even in complete darkness (Dieter, Hu, Knill, Blake, & Tadin, 2014). The duration of sounds can bias the perceived duration of concurrent visual stimuli (Romei, de Haas, Mok, & Driver, 2011) and sensitivity for a brief flash increases parametrically with the duration of a co-occurring sound (de Haas, Cecere, Cullen, Driver, & Romei, 2013). The noise level of visual stimulus representations in retinotopic cortex is affected by the (in)congruency of co-occurring sounds (de Haas, Schwarzkopf, Urner, & Rees, 2013). Category-specific sounds and visual imagery can be decoded from early visual cortex, even with eyes closed (Vetter, Smith, & Muckli, 2014) and the same is true for imagined hand actions (Pilgramm et al., 2015). Vice versa, the location of visual stimuli can bias the perceived origin of sounds (Thomas, 1941) and a visible face articulating a syllable can bias the perception of a concurrently presented (different) syllable (McGurk & MacDonald, 1976).

Firestone and Scholl argue that multisensory effects can be reconciled with total encapsulation. Their inflexible nature and short latency would provide evidence they happen ‘within perception itself’, rather than reflecting the effect of ‘more central cognitive processes on perception’. However, multisensory effects have different temporal latencies and occur at multiple different levels of processing, from direct cross-talk between primary sensory areas to top-down feedback from ‘association cortex’ (de Haas & Rees, 2010; Driver & Noesselt, 2008). They may further be subject to attentional (Navarra, Alsius, Soto-Faraco, & Spence, 2010), motivational (Bruns, Maiworm, & Röder, 2014) and expectation based (Gau & Noppeney, 2015) modulations themselves. Therefore, evidence regarding a strictly horizontal nature of multisensory effects seems ambiguous at best. If total encapsulation hinges on this hypothesis, it needs to be clearer. Specifically, what type of neural or behavioural evidence could refute it?

A second – perhaps more definitive – counterexample is *attentional* modulation of vision. Firestone and Scholl acknowledge that attention can change what we see (cf. Anton-Erxleben, Abrams, & Carrasco, 2011; Carrasco, Fuller, & Ling, 2008) and that these effects can be under intentional control. For instance, voluntary attention can induce changes in the perceived spatial frequency (Abrams, Barbot, & Carrasco, 2010), contrast (Liu, Abrams, & Carrasco, 2009) and position (Suzuki & Cavanagh, 1997) of visual stimuli. Withdrawal of attention can induce perceptual blur (Montagna, Pestilli, & Carrasco, 2009) and reduce visual sensitivity (Carmel, Thorne, Rees, & Lavie, 2011) and sensory adaptation (Rees, Frith, & Lavie, 1997). Nevertheless, Firestone and Scholl argue for total encapsulation. On such an account, attention would not interfere with visual processing per se but with the *input* to this process, ‘similar to changing what we see by moving our eyes’ or ‘turning off the lights’.
Attention-related spatial distortions and changes in acuity have been linked to effects on the spatial tuning of visual neurons (Anton-Erxleben & Carrasco, 2013; Baruch & Yeshurun, 2014). Receptive fields can shift and grow towards, or shrink around attended targets (e.g., Womelsdorf, Anton-Erxleben, & Treue, 2008). Such effects go beyond mere amplitude modulation and can provide important evidence regarding their locus. In a recent study (de Haas, Schwarzkopf, Anderson, & Rees, 2014), we investigated the effects of attentional load at fixation on neuronal spatial tuning in early visual cortices. Participants performed either a hard or easy fixation task while retinotopic mapping stimuli traversed the surrounding visual field. Importantly, stimuli were identical in both conditions – only the task instructions differed. When participants performed the harder task, and consequently could pay less attention to the task-irrelevant mapping stimuli (Lavie, 2005), this yielded a blurrier neural representation of the surround, as well as a centrifugal repulsion of population receptive fields in V1-3 (pRFs; Dumoulin & Wandell, 2008). Importantly, this repulsion in V1-3 was accompanied by a centripetal attraction of pRFs in the intraparietal sulcus (IPS), perhaps because the larger receptive fields in IPS specifically encode the attended location (Klein, Harvey, & Dumoulin, 2014). Critically, retinotopic shifts merely inherited from input modulations cannot trivially explain such opposing shifts because any such effect should be the same (or very similar) across the visual processing hierarchy.

How can one reconcile these findings with total encapsulation? We can only think of one way: redefining ‘visual processing’ in a way that excludes processing associated with retinotopic tuning of visual cortex but includes feedback processes from multisensory areas (see above). This seems hard to reconcile with the widespread evidence that visually tuned neuronal populations in occipital cortex are involved with visual processing. We instead argue that attentional and multisensory modulations are inconsistent with total encapsulation and at least here the line between cognition and perception is blurred. Firestone and Scholl concede that accepting this would be far less revolutionary than many of the claims they attack. We second their demand to back up extraordinary claims with rigorous evidence and applaud the standards they propose. Many effects they discuss may indeed fail to meet these standards. But precisely because attentional and multisensory effects are well established, total encapsulation itself strikes us as an extraordinary claim that is not supported by the available evidence.

Acknowledgments

This work was supported by a research fellowship from the Deutsche Forschungsgemeinschaft (BdH; HA 7574/1-1), European Research Council Starting Grant 310829 (DSS, BdH) and the Wellcome Trust (GR).

References


