



Attention and cognitive penetration: reflections on Dustin Stokes' *Thinking and Perceiving*

Wayne Wu¹

Accepted: 15 July 2023
© The Author(s) 2023

Abstract

Dustin Stokes book *Thinking and Perceiving* is a substantial achievement. In this comment, I discuss issues related to cognitive penetration. While I agree with Stokes' criticisms of Fodor and Pylyshyn's discussion of cognitive penetration with respect to the role of attention, I provide a supporting, but different argument against how they understand attention. I also emphasize that the common appeal to behavioural data in arguing for cognitive penetration is less effective than an argument that supplements behavioural data with computational models. I sketch how drawing on computational models strengthens the case for the cognitive penetration of attention by intention. This suggests that cognitive penetration is as common as intentional action.

Keywords Attention · Cognitive penetration · Computational model

Dustin Stokes has written a terrific book that provides an excellent, opinionated, and illuminating discussion of the relation between thought and perception, with special emphasis on cognitive penetration and perceptual expertise. Theorists new to the debate and old hands will benefit from Stokes' rich and important discussion.

Stokes emphasizes a novel perspective on cognitive penetration: we should care about cognitive penetration of perception by assessing its specific consequences. This involves showing that cognition affects perception and that the effects are *consequential* in ways that matter to us. Among relevant consequences are the broad range of epistemological upshots that arise if thought changes perception such as if *believing is seeing*. If we see what we believe, vision's content is inherited from belief. Such

✉ Wayne Wu
waynewu@andrew.cmu.edu

¹ Department of Philosophy and Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA

inheritance would undercut vision's role as a neutral tribunal for belief, a consequential upshot indeed!

Stokes contrasts two different processes in respect of cognition's influence on attention and thereby on vision. That first, (a), draws on an old perspective of Jerry Fodor and Zenon Pylyshyn's:

(a) Cognitive State → (Spatial) Attention Shift → Perceptual Experience.

The second is Stokes' variation on the basic flow of information/causation.

(b) Cognitive State → Non-agential Selective (feature/object) Attention → Perceptual Experience.

The first apparently disallows cognitive penetration, the latter allows for it. The two processes are construed contrastively, but do they really contrast?

Consider causal chain (a). A cognitive state can affect spatial attention, as when an intention to attend moves the eye to the left, namely switches *overt* visual attention's target. It changes what you look at. Switching your intention to then attend to the right shifts spatial attention, and the eye, to the right. Cognition directs attention. Now add to this uncontested fact about *top-down* attention a further widely endorsed thesis: *attention gates consciousness*. On this view, one is visually conscious of only what one is attending to (this claim requires some finessing, see (Wu, 2014, Ch. 5)). Note that with these two widely endorsed claims about attention, we have an uncontroversial cognitive effect on attention and a substantial consequence on experience. A failure to attend to some object—oh....I don't know, maybe a gorilla (Simons & Chabris, 1999)?—means that one fails to be conscious of it and hence that one is not in a position to respond to it in normal ways. Thus, one would not be able to report on the gorilla, to form beliefs about it, or to otherwise respond to it. The same argument can be made without eye movement, so with covert attention, hence without requiring that the retinal input changes (see experiments by (Mack and Rock 1998)).

If we opt for a simple consequentialist view, then the debate about cognitive penetration was resolved a long time ago, and apparently without most people noticing it! We need only recognize an uncontroversial cognitive effect on visual attention that drastically alters visual experience, namely one that leads to inattentional *blindness* (Mack and Rock 1998). Where attention is focused has clear downstream consequences as accidents due to texting while driving make clear. This argument seems *too* easy especially since cognitive penetration is bitterly contested. Further, the move just made flanks the productive and expansive appeal that Stokes makes of detailed empirical data to settle the debate after years of wrangling. The question I am raising is whether consequentialism might be too permissive a condition for establishing cognitive penetration.

But weren't Fodor and Pylyshyn right that mechanism (a) *closes off* cognitive penetration? I have never understood this claim, and it seems to me confused about attention. If spatial attention is just an eye movement, so overt attention, then the influence of cognition is on bodily action and not on perception. Stokes was one of the first philosophers, I think, to state this clearly. Cognition's influence is indirect

for it moves the eye which changes the visual input into the perceptual system and thus, changes one's experience. Attention affects inputs into the visual system. So far, so clear. However, the relevant form of attention we should be considering is *covert* visual attention, attention without eye movement. Here, it seems bizarre to say that covert attention changes the input into the sensory system in the same way that moving the eye does. For example, if you think that covert visual attention is a spotlight-like mechanism, then where does the spotlight operate *before* the visual system gets involved?

Certainly not before the eye. That would be a horrible confusion. Maybe at the eye, such as the retina? Well there's no evidence for that as far as I know. There is evidence that attentional modulations can occur as early as the lateral geniculate nucleus (LGN), but that structure is arguably part of the visual system and visual processing divisions therein ramify into cortical visual processing. Let's take the LGN as the earliest point where neural changes correlated with covert attention engage. It is not then true that covert attention affects inputs into the visual system. Rather, it acts on processing *within* the visual system.

The Fodor/Pylyshyn camp has not shown that mechanism (a) closes off cognitive penetration in covert attention. If so, then Stokes does not *need* to shift discussion to nonspatial forms of attention, as he does in the book, in his resistance to the Fodor/Pylyshyn view. This would be to concede too much. Stokes' appeal to nonspatial attention has given us a *further* reason to reject the Fodor/Pylyshyn position on attention, but we already had reason to reject it.

One thing that one might say in response is that the issue isn't visual information processing but visual *experience*. I agree that cognitive effects on experience are harder to demonstrate, but for interesting reasons. Here, I find Stokes' discussion of perceptual expertise very illuminating and I urge folks to read those chapters and learn from them. I will say that on a consequentialist view, as I understand it, if cognition were, in affecting attention, to then affect visual processing so as to influence *behavior*, this would be consequential. I think this happens all the time, so let me elaborate that.

My view is that visual covert attention supervenes on appropriate changes in visual processing (Wu 2023, chap. 2). In the biased competition account of visual attention, the relevant changes to visual processing are via cognitive representations such as task representations in working memory (this being realized, in part, in prefrontal cortical activity). So, the key idea is that attention is realized in changes to visual processing to yield visual attention to relevant targets, even if the eye doesn't move. On this view, visual attention is not an input into visual processing. It *emerges* from selective visual processing (cf. (Desimone and Duncan 1995; Hommel et al., 2019)).

This leads to my second point, a comment on computation. When I started thinking about cognitive penetration of vision as an *empirical* thesis, indeed as a biological thesis, that is as a thesis about the primate visual system, it struck me that what was needed in the discussion is an account of penetration that draws on multiple levels of analysis. I would say that in the renewing of the debate about cognitive penetration by Siegel (2012), Macpherson (2012) and Stokes (2012) among others, primary emphasis was placed on behavioral data. Behavioral data is important, a basic starting point for theorizing in empirical cognitive science. Yet I felt that drawing on just

this data could only get you so far in adjudicating the issue. In all the experiments examined, the data is friendly to proponents and opponents of cognitive penetration. Each can construct ways to explain the data given underdetermination.

Let me elaborate this concern with respect to the Bruner and Goodman (1947) experiment that Stokes discusses (Stokes op. cit.). In this study, the experimenters report a difference between how wealthy and poor kids indicate the apparent size of coins. One might conclude that economic valuation penetrates visual experience of coin size. An alternative account is that the reports reflect a *memory* effect. In indicating the perceived size of the disks or coins, subjects had to look back and forth between the target disk/coin and a projected light circle whose size they adjusted to report apparent size. Comparison across fixations requires visual short term memory, yet memory and report are very much subject to cognitive penetration. The adjustment kids make of the light circle to report apparent size is based on visual short term memory of the size of the coin...and potentially other knowledge they have. Since the kids in the two groups have different experiences, expectations and values, these informational factors could just as well affect their memory rather than their initial experience, with these differences explaining variation in their reports. I'm not arguing that this is the correct explanation, but demonstrating that alternative explanations are always available in explaining the behavioral data.

What about adding neuroscience? As neuroscientists will point out, the visual system is massively modulated top-down, and so the circuitry seems to be built for informational penetration. Yet the issue of cognitive penetration will not be settled by anatomical considerations alone nor simply by adding information about brain activity. I have emphasized that we should remember that the empirical thesis about cognitive penetration is at its core a *computational claim regarding informational processing* (Wu 2013; 2017). So, cognitive penetration is a specific hypothesis in cognitive science. It is the claim that information from cognition plays a specific role in visual information processing. We have to make sense of the idea that visual processing computes over cognitive content.

Neural anatomy, neural activity, and behavioral data will not be enough to establish a *computational* thesis. We also need computational models. Specifically, we should look at the best computational models of visual processing, fit that to cognitive penetration as a computational claim in respect of performance on a specific task, and then look for behavioral and neural evidence linked by a model. This is the approach I took in my (2017) article that Stokes mentions, an article I am sad to say is difficult to read, so let me try to do a better job.

Begin with some behavior, namely the engagement of attention in performing a task, say one that requires some response to one of two Gabor patches (contrast gradients), left and right of fixation. For example, you might have to report the orientation of that Gabor, say a leftward versus rightward tilt. For me, it is enough that the subject is able to perform a task selectively to show that the subject's selective attention is engaged. Since such selective visual attention in task performance is often associated with distinctive neural changes, from the shrinking of receptive fields around task relevant objects to the amplification or sharpening of signals carrying stimulus information, all in the visual system, we have correlates of neural selection that paral-

lel behavioral selection. None of this, however, yet shows that vision is computing over cognitive contents.

This is where a computational model is critical: it bridges neural-behavioral data to the computational thesis of cognitive penetration. Consider the *divisive normalization* model from David Heeger and John Reynolds (2009) that explains why populations of neurons change their response in a way that is biased towards the task relevant object, here the target on the right that is cued:

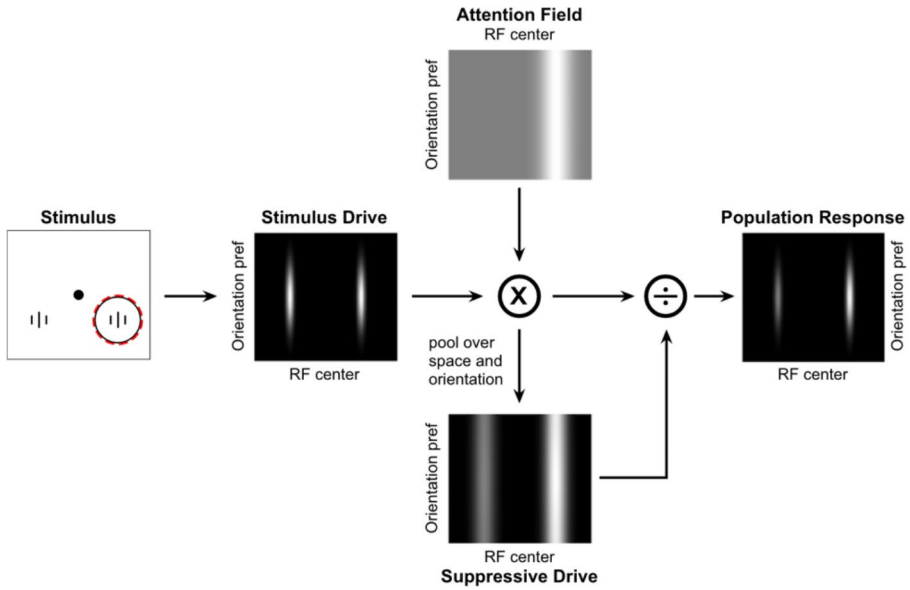


Fig. 1 Reynolds and Heeger divisive normalization model of attention. Reprinted from *Neuron*, 61, John Reynolds and David Heeger, “The Normalization Model of Attention”, pp. 168–84, 2009, with permission from Elsevier

The fundamental idea is normalization, division. Consider the input (Stimulus) into the visual system. The aim is to model the system’s output response given that input. The first step is the multiplication of a representation of the stimuli, the *Stimulus Drive*, understood as indicating the neural responses of visual neurons to two stimuli, one on the left, the other on the right of the visual field, by a second representation, the *Attention Field*. Think of each point in the bright vertical bars in the Stimulus Drive as giving the magnitude of the response of neurons whose receptive fields contain one of the visible stimuli (left or right). The X-axis gives the location of the neuron’s receptive field in the x-dimension, the Y-axis indicates what orientation the neuron prefers, and the level of brightness indicates the strength of the response (dark areas mean no response by the neurons represented). Divisive normalization (division) outputs the overall *Population Response*, a difference in the response that shows a bias towards one of the two stimuli, here the task relevant stimulus on the right. The key point is that the selectivity that we observe in behavior predicated on visual attention has a correlate, and presumably a basis, in the selective shifts in neural processing as modeled by divisive normalization.

We now have materials to make a stronger inference to cognitive penetration. Let's return to the *Attention Field*. What the *Attention Field* highlights is the *task relevant location of the target* namely its location on the right. Yet task relevance is part of how the subject is instructed or trained, and so part of the subject's intention understood as a cognitive state. That intention, encoding the task relevant target, *must* make a difference to visual processing in order to produce the correct behavior, namely behavior directed to the right. How else is that behavior consistently directed in the *correct* way, that is, directed as intended? If so, then I think the way to think about the Attention Field is that it carries the content of the agent's intention, here the location of the target on the right. But then, the circuit shows us a plausible case of cognitive penetration where it looks like visual information processing is computing over a cognitive representation of the spatial target of the intention.

Finally, if the shift in visual information signaling facilitates behavior, then the cognitive penetration in divisive normalization has substantial consequences: it enables us to be effective visual agents. I think this is the most complete argument one can make in favor of cognitive penetration. It unifies the behavioral and neural data through a computational model, which is appropriate since cognitive penetration is a computational thesis. Further, the model fleshes out the causal chain (a) above. Similar ideas can also flesh out Stokes' (b). So, if both (a) and (b) gives us consequential cognitive penetration, then Stokes and I can both be satisfied.

Funding Open Access funding provided by Carnegie Mellon University.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bruner, J. S., & Cecile, C. G. (1947). Value and need as Organizing factors in perception. *The Journal of Abnormal and Social Psychology*, 42(1), 33–44.
- Desimone, R., and John Duncan (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18(1), 193–222. <https://doi.org/10.1146/annurev.ne.18.030195.001205>.
- Hommel, B., Chapman, C. S., Cisek, P., Neyedli, H. F., Song, J. H., & Timothy, N. W. (2019). No one knows what attention is. *Attention Perception & Psychophysics*, 81(7), 2288–2303.
- Mack, A., and Irvin Rock (1998). *Inattention blindness*. Cambridge, MA: MIT Press.
- Macpherson, F. (2012). Cognitive penetration of Colour Experience: Rethinking the issue in light of an indirect mechanism. *Philosophy and Phenomenological Research*, 84(1), 24–62.
- Reynolds, J. H., and David J. Heeger (2009). The normalization model of attention. *Neuron*, 61(2), 168–185. <https://doi.org/10.1016/j.neuron.2009.01.002>.
- Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs*, 46(2), 201–222.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained Inattention blindness for dynamic events. *Perception*, 28(9), 1059–1074. <https://doi.org/10.1068/p281059>.

- Stokes, D. (2012). Perceiving and desiring: A New look at the cognitive penetrability of experience. *Philosophical Studies*, 158(3), 479–492.
- Wu, W. (2013). Visual spatial constancy and modularity: Does intention penetrate vision? *Philosophical Studies*, 165(2):647–69. <https://doi.org/10.1007/s11098-012-9971-y>.
- Wu, W. (2014). *Attention*. Abingdon, UK: Routledge.
- Wu, W. (2017). Shaking up the mind's ground floor: The cognitive penetration of visual attention. *Journal of Philosophy*, 114(1):5–32.
- Wu, W. (2023). *Movements of the mind*. Oxford: Oxford University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.