The studies purporting to control for demand were themselves demand-ridden (e.g., Durgin et al. 2012). Participants were warned not to respond the way others did, inflating hill slant estimates to please experimenters. Not surprisingly, those participants responded by giving low estimates of slant compared with the estimates of unwarned participants. Far from eliminating implicit demand, such instructions introduce explicit demand. The literature on "debiasing" (e.g., Schwarz 2015) indicates that such warnings tend to bias, rather than debias, judgments.

Participants in Durgin et al. (2009) were told that the backpack contained equipment to measure muscle potential. An elaborate story ensured that participants were alerted to the irrelevance of their experience of the backpack for judging slant. Additionally, a noisy cooling fan on the backpack served as a constant reminder of its irrelevance. Rather than eliminating demand, the elaborate effort to make the backpack both salient and distinct ensured that its heaviness would be segmented from other bioennergetic cues when making slant estimates. The results seem foreordained by the procedures. The findings likely reflect not only the odd backpack manipulation, but also the fact that the "hill" was only a 2-meter-long ramp. This hill did not afford the opportunity to walk more than a step or two, rendering any bioenergetic costs of wearing a backpack irrelevant (Proffitt 2009).

To assess awareness, experimenters (e.g., Durgin et al. 2009) first asked participants to consider the backpack heaviness and then asked for their hypotheses. The literature on assessing awareness has long warned against such procedures (e.g., Bargh & Chartrand, 2000; Dulany 1962), because participants cannot reliably distinguish hypotheses elicited by being asked questions from hypotheses actively generated during the experiment. To avoid such contamination, recommended methods involve carefully designed funnel interviews (e.g., Dulany 1962), which were not used in this research.

Other studies have found that replenishing depleted glucose can lower slant estimates (Schnall et al. 2010), but F&S suggest this occurs not because changes in resources affect perception, but because added glucose empowers participants to resist experimenter demands. We are not aware of any data supporting glucose effects on susceptibility to demand. More important, such suggestions cannot account for the results of multiple new studies, only some of which we cited earlier, which are simultaneously immune to the criticisms of F&S and supportive of the perceptual effects they attack.

ACKNOWLEDGMENTS
This work was partially supported by grant number BCS-1252079 from the National Science Foundation to Clore.

Bottoms up! How top-down pitfalls ensnare speech perception researchers, too

doi:10.1017/S0140525X15002745, e236

Anne Cutler[a] and Dennis Norris[b]

[a]*The MARCS Institute, Western Sydney University, Locked Bag 1797, Penrith South, NSW 2751, Australia;* [b]*MRC Cognition and Brain Sciences Unit, Cambridge CB2 2EF, United Kingdom.*

a.cutler@westernsydney.edu.au dennis.norris@mrc-cbu.cam.ac.uk
www.westernsydney.edu.au/marcs/our_team/researchers/professor_anne_cutler
www.mrc-cbu.cam.ac.uk/people/dennis.norris

Abstract: Not only can the pitfalls that Firestone & Scholl (F&S) identify be generalised across multiple studies within the field of visual perception, but also they have general application outside the field wherever perceptual and cognitive processing are compared. We call attention to the widespread susceptibility of research on the perception of speech to versions of the same pitfalls.

Firestone & Scholl (F&S) review an extensive body of research on visual perception. Claims of higher-level effects on lower-level processes, they show, have swept over this research field like a "tidal wave." Unsurprisingly, other areas of cognitive psychology have been similarly inundated.

Auditory perception and visual perception are alike in the questions they raise about the interplay of cognitive processing with sensory and perceptual analysis of a highly complex and externally determined input; and like visual perception, the processes underlying auditory perception have been well mapped in recent decades. Many of the features of the vision literature F&S note have direct auditory counterparts, such as highly intuitive demonstrations (the McGurk effect, whereby auditory input of [b] combined with visual articulation of [g] produces a percept of [d]; McGurk & MacDonald 1976), control of peripheral attention shift (the "cocktail-party phenomenon," attending to one interlocutor in a crowd of talking people), or novel pop-out effects, for example, for certain phonologically illegal sequences (Weber 2001).

Auditory and visual perception differ, however, not only in sensory modality but also in the input domain: Visual signals play out in space; auditory input arrives across time. The temporal input dimension has had implications for how the equivalent debate in the speech perception literature has played out; it has proved natural and compelling to treat the question of modularity as one concerning the temporal order of processing – has the bottom-up processing order (e.g., of speech sounds before the words they occur in) been compromised? Studies in which ambiguous speech sounds are categorised differently in varying lexically biased contexts (e.g., a [d/t] ambiguity reported as "d" before *-eep* but as "t" before *-eek* because *deep* and *teak* are words but *teep* and *deek* are not; Ganong 1980) were initially taken as evidence for top-down effects. (Note, however, that rather than tapping directly into perceptual processes, categorisation tasks may largely reflect metalinguistic judgments, as per F&S's Pitfall 2.) This work prompted follow-ups showing, for example, stronger lexical effects in slower responses (Fox 1984), and no build-up of effects with an ambiguous sound in syllable-final rather than syllable-initial position (McQueen 1991); these temporal arguments suggested a response bias account (F&S's Pitfall 3).

In general, F&S's Pitfall 1 (a confirmatory approach) has been the hallmark of much of the pro-top-down speech perception literature. Most of that literature takes the form of a catalogue of findings that are consistent with top-down effects but are not diagnostic. There is frequently little evidence that alternative feed-forward explanations have been considered. One of the few exceptions comes from the study of compensation for coarticulation, a known low-level process in speech perception whereby cues to phonetic contrasts may be weighted differently depending on immediately preceding sounds. An influential study by Elman and McClelland (1988) reported that interpretation of a constant word-initial [t/k] ambiguity could be affected by the lexically determined interpretation of a constant immediately preceding word-final [s/sh] ambiguity (whether it served as the final sound of *Christmas* or *foolish*). Pitt and McQueen (1998) reasoned that if this compensation was a necessary consequence of the lexical effect (rather than of transitional probability, as they argued; that is, an artefact as per F&S's Pitfall 4), then if there is no compensation effect there should be no lexical effect either. With the [s/sh] occurring instead in words balanced for transitional probability (*juice*, *bush*), the word-initial [t/k] compensation disappeared; but the lexical [s/sh] effect remained. Such studies (testing disconfirmation predictions) are, however, as in the vision literature, vanishingly rare.

In our account in this journal of these and similar sets of studies (Norris et al. 2000), we concluded, as F&S do for the vision literature, that there was then no viable evidence for top-down penetrability of speech-perception processes. In a more recent review (Norris et al. 2016) we reach the same conclusion regarding current research programs in which similar claims have been reworded in terms of prediction processes ("predictive coding").

Priming effects (F&S's Pitfall 6) are more or less the bread and butter of spoken-word recognition research, so that psychological studies tend to preserve the memory/perception distinction; but in an essentially separate line of speech perception studies, from the branch of linguistics known as sociophonetics, the distinction has in our opinion been blurred. Typical results from this literature include listeners' matching of heard tokens to synthesised vowel comparison tokens being influenced by (a) telling participants the speaker was from Detroit versus Canada (Niedzielski 1999), (b) labeling participants' response sheets "Australian" versus "New Zealander" (Hay et al. 2006), or (c) having a stuffed toy kangaroo or koala versus a stuffed toy kiwi in the room (Hay & Drager 2010). In fairness to these authors, we note that they do not propound large claims concerning penetration of cognition into primary auditory processing (they interpret their results in terms of reference to listening experience). It seems to us, however, that a rich trove of possible new findings could appear if researchers would adopt F&S's advice and debrief participants, then correlate the match responses to debriefing outcomes.

A comprehensive and thorough review of a substantial body of research (with potentially important implications for theory) is always a great help to researchers – and especially useful if it uncovers new patterns such as, in this case, a systematic set of deficiencies. But in the present article, a service has been performed for researchers of the future as well, in the form of a checklist against which the evidence for theoretical claims can be evaluated. Only research reports that pass (or at least explicitly address) F&S's six criteria can henceforth become part of the serious theoretical conversation. As we have indicated, these criteria have application beyond visual perception; at least speech perception can use them too. Thus, we salute F&S for performing a signal service to the cognitive psychology community. Bottoms up!

Attention and multisensory modulation argue against total encapsulation

doi:10.1017/S0140525X1500254X, e237

Benjamin de Haas,[a,b] Dietrich Samuel Schwarzkopf,[a,b] and Geraint Rees[a,c]

[a]*Institute of Cognitive Neuroscience, University College London, WC1N 3AR London, United Kingdom;* [b]*Experimental Psychology, University College London, WC1H 0AP London, United Kingdom;* [c]*Wellcome Trust Centre for Neuroimaging, University College London, WC1N 3BG London, United Kingdom.*

benjamin.haas.09@ucl.ac.uk s.schwarzkopf@ucl.ac.uk g.rees@ucl.ac.uk
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 & Scholl (F&S) postulate that vision proceeds without *any* direct interference from cognition. We argue that this view is extreme and not in line with the available evidence. Specifically, we discuss two well-established counterexamples: 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."

Firestone & Scholl (F&S) 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, p. 66). Their hypothesis is much more ambitious: "perception proceeds without any direct, unmediated influence from cognition" (sect. 5, para. 1). 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 reappearance of visual flashes (Shams et al. 2000). Such reappearances increase objective sensitivity for visual features of the flash (Berger et al. 2003) and are linked to individual structure and function of primary visual cortex (de Haas et al. 2012; Watkins et al. 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 et al. 2014). The duration of sounds can bias the perceived duration of concurrent visual stimuli (Romei et al. 2011), and sensitivity for a brief flash increases parametrically with the duration of a co-occurring sound (de Haas et al. 2013a). The noise level of visual stimulus representations in retinotopic cortex is affected by the (in)congruency of co-occurring sounds (de Haas et al. 2013b). Category-specific sounds and visual imagery can be decoded from early visual cortex, even with eyes closed (Vetter et al. 2014), and the same is true for imagined hand actions (Pilgramm et al. 2016). At the same time, 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). F&S argue that multisensory effects can be reconciled with total encapsulation. The inflexible nature and short latency of such effects would provide evidence they happen "within perception itself," rather than reflecting the effect of "more central cognitive processes *on* perception" (sect. 2.4, para. 1). However, multisensory effects have different temporal latencies and occur at multiple 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 et al. 2010), motivational (Bruns et al. 2014), and expectation-based (Gau & Noppeney 2015) modulations. Therefore, evidence regarding a strictly horizontal nature of multisensory effects seems ambiguous at best. If total encapsulation hinges on the hypothesis of strictly horizontal effects, this hypothesis 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. F&S acknowledge that attention can change what we see (cf. Anton-Erxleben et al. 2011; Carrasco et al. 2008) and that these effects can be under intentional control. For example, voluntary attention can induce changes in the perceived spatial frequency (Abrams et al. 2010), contrast (Liu et al. 2009), and position (Suzuki & Cavanagh 1997) of visual stimuli. Withdrawal of attention can induce perceptual blur (Montagna et al. 2009) and reduce visual sensitivity (Carmel et al. 2011) and sensory adaptation (Rees et al. 1997). Nevertheless, F&S 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 the lights off" (sect. 4.5, para. 4).

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 et al. 2008). Such effects go beyond mere amplitude modulation and can provide important evidence regarding their locus. In a recent study (de Haas et al. 2014), we investigated the effects of attentional load at fixation on neuronal spatial tuning in early visual cortices. Participants performed either a hard or an easy fixation task while retinotopic mapping stimuli traversed the surrounding visual field. Importantly, stimuli were identical in both conditions – only the task instructions differed. Performing the harder task, and consequently