Available online at www.sciencedirect.com

ScienceDirect

Physics of Life Reviews 7 (2010) 295–298

PHYSICS of LIFE reviews

www.elsevier.com/locate/plrev

Reply to comment

Bayesian priors and multisensory integration at multiple levels of visual processing
Reply to comments on "Crossmodal influences on visual perception"

Ladan Shams [a,b,c,*], Robyn Kim [a]

[a] Department of Psychology, University of California, Los Angeles, United States
[b] Department of Biomedical Engineering, University of California, Los Angeles, United States
[c] Interdepartmental Neuroscience Program, University of California, Los Angeles, United States

Received 9 July 2010; accepted 28 July 2010

Available online 1 August 2010

Communicated by L. Perlovsky

Keywords: Bayesian inference; Bayesian priors; Multisensory integration; Visual perception; Causal inference; Crossmodal binding

---

We would like to thank all commentators for their insightful and thought-provoking commentaries. We find it gratifying that the commentators represent a diverse array of expertise, as they have enriched the discussion with their different perspectives on this topic. The commentaries have also identified and highlighted important open questions in this field; an important step in advancing the field. Here, we discuss some of the important theoretical issues and observations raised by de Haas and Rees [1], Spence [2], Alais [3], Vroomen [4], and Barone [5].

de Haas and Rees [1] underscore that crossmodal interactions appear to occur at multiple levels of processing, and argue that any theory of multisensory perception should account for this phenomenon. We agree with this assessment. As discussed in the target paper [6], crossmodal interactions have been reported for a variety of visual tasks ranging from low-level perceptual tasks such as detection and motion perception, to high-level tasks such as object recognition. Even for a simple low-level task, there appear to be interactions between modalities at multiple levels of processing. For example, as noted by de Haas and Rees [1], in the numerosity judgment task (discussed in Sections 3 and 4 of the target paper), fMRI revealed interactions in areas ranging from superior colliculus, V1, and STS [7,8], and MEG revealed early-onset interactions in occipital regions followed by later interactions in parietal and frontal areas [9]. Interactions at multiple levels of processing are indeed consistent with a Bayesian inference scheme in which both likelihoods (sensory representations) and priors (expectations) play a role in the perceptual process. While the combination of sensory information (likelihood functions) can occur at an early neural level of processing (e.g., V1, superior colliculus, or thalamus), the priors may involve interactions at a range of neural processing levels. For example, in

---

DOIs of commentaries: 10.1016/j.plrev.2010.06.010, 10.1016/j.plrev.2010.07.002, 10.1016/j.plrev.2010.06.004, 10.1016/j.plrev.2010.06.007, 10.1016/j.plrev.2010.07.003.

[*] Corresponding author at: 7445 Franz Hall, Department of Psychology, University of California, Los Angeles, 90095-1563, United States. Tel.: +1 310 206 3630; fax: +1 310 267 2141.

E-mail address: ladan@psych.ucla.edu (L. Shams).

1571-0645/$ – see front matter
doi:10.1016/j.plrev.2010.07.006

a spatial localization task, there are two components to priors involved in the inference process [10]. One component encodes the expectation about the spatial location of objects. This prior likely reflects the life-long experience of the observer, capturing the statistics of spatial layout of objects relative to the observer. Another component of priors is the expectation of the auditory and visual stimuli having a common cause. This prior would also represent the statistics of the environment (how often simultaneous visual and auditory stimuli have originated from the same object in the past experience). Both of these priors can be learned by experience or hard-wired, and in either case, they could be implemented at a low level of neural processing. On the other hand, both of these priors can also be influenced by high-level knowledge, e.g., by instructions given by the experimenter about the setup of the stimuli, or by indirect clues about the location of objects or unity of the stimuli (see [11] for an example). This type of prior or modulation of the prior must involve top–down pathways and higher levels of neural processing. While these two kinds of priors have different origins and may involve different neural pathways, they both have the same kind of influence on the perceptual inference process, and their effects may not be distinguishable behaviorally. Therefore, our answer to de Haas and Rees's question of "Can semantic instructions about modality or feature reliability alter Bayesian priors for modalities or features?" is yes, we expect that at least some types of instructions (e.g., about the causal structure of the stimuli) can change the integration process by altering priors. Indeed the example Vroomen [4] gives regarding the effect of telling participants about the nature of sounds (that they are speech sounds) and its consequence on the integration of auditory-visual information and speech perception is a very good example of the influence of instructions on modifying the perceptual process through a change in priors. Less clear is whether instructions about feature reliabilities can affect the weight that is assigned to the different modalities.

de Haas and Rees [1] also ask the excellent question: "How would such an effect differ from the same kind of effect mediated by implicit learning?" In the Bayesian decision theory framework discussed in the target paper, there is no difference between the two types of priors (high-level and low-level) from a computational point of view. In other words, both types of priors would have the same behavioral consequence. The mechanisms of these influences are currently entirely unknown. It remains for the future research to shed light on how the different kinds of priors get combined and exert influence in the inference process, what the underlying neural pathways for the top–down and bottom–up priors are, to what extent low-level priors (representing statistics of the environment acquired by learning or hard-wired) can be modified by top–down knowledge (e.g., instructions), and so on. These are all fundamental and challenging questions facing the field of perception as a whole.

The extant data suggests that similar computational principles govern integration for tasks involving different domains and levels of processing. Vroomen [4] draws attention to parallels between visual and auditory perception. Vroomen [4] and Spence [2] both point out studies that—although they use very different experimental paradigms—both suggest causal inference also plays a role in auditory-visual speech perception. Consistent with these observations, in a forthcoming review paper [12], we argue that causal inference is a process that is at the core of perception, it occurs in both unisensory and multisensory perceptual tasks, and the same type of computation appears to be employed by the perceptual system in a variety of domains.

Spence [2] underscores the importance of binding problem and causal inference in multisensory perception when there are multiple sensory stimuli. While we completely agree that this problem is significantly more critical and challenging in conditions with multiple sensory signals (i.e., in natural conditions), we would like to point out that this problem still exists even in the simplest scenario, in which there are only two sensory signals, e.g., a visual signal and an auditory signal. Even in this simple case, there are two possible causal structures that could have given rise to the signals (common cause, and independent causes), as depicted in Fig. 5c of target article. Indeed the traditional studies of crossmodal interactions had bypassed the problem of causal inference by always using stimuli with small to moderate conflict, and thus remaining within the realm of inferred common cause that results in a fusion of signals.

Alais [3] interprets a recent finding about influence of synchronous but spatially non-informative sound on visual spatial target detection (also see last paragraph of Section 2.1 of target article) in light of the binding problem. Alais makes the point that in the absence of any spatial auditory information, the temporal consistency between the auditory and visual signals has to be very strong and compelling to lead to binding of the two, otherwise it could potentially result in binding of wrong signals together. This is a very interesting observation, and is exactly what would be predicted by an observer performing Bayesian causal inference [10,12,13] (also see Section 7 of target article). In making an inference about whether two signals were caused by the same source or not, the Bayesian observer would take evidence from various features into account. The consistency in time, space, structure (if the stimuli contain structure), and semantics (if the stimuli have semantic content) would collectively contribute to the inference of a common cause [10]. The lower the degree of consistency in any of these dimensions, the lower the probability of a common cause inferred by the observer. Therefore, a similar probability of a common cause (in the inference process) can be achieved by having moderate consistency across all features, or having a strong degree of consistency in one feature and lack of information in others.

de Haas and Rees [1] ask a novel question: "... could multimodal cues be combined between different individuals in a Bayes-optimal fashion?" For example, if a blind person and a deaf person cooperate, can they jointly achieve perceptual estimates that are as statistically optimal as those of a sighted-hearing person? This is an interesting question because it highlights important issues about perceptual processing that are not well-understood. We believe that a near-optimal perception for the team could be possible depending on whether or not the two individuals have the same loss/utility function for the given task, and on the degree of conscious access the individuals have to the posterior distributions underlying their sensory estimates, and the degree that they are able to communicate this information. Contrary to the common belief that low-level perception is highly optimized by evolution and therefore the loss function used for these tasks is uniform across observers, we have recently found that there is variability across individuals in the perceptual decision strategy (and hence the loss function) even in a low-level perceptual task [14].

Barone [5] offers relevant information about the cross-connectivity between sensory brain areas. We agree with Barone that recent studies have not provided convincing evidence of visual areas being activated (as opposed to modulated) by non-visual stimuli. The reports of response in Brodmann areas 17, 18, and 19 to acoustic stimuli in older literature [15,16] have indeed been in cat, and Barone points out that the primate visual cortex may differ from those of other mammals in this respect. Indeed we agree that it is important to note that the existence of multisensory integration in a given area (e.g., V1) does not necessarily imply that the area can be designated as a multisensory area *per se*.

Spence [2] raises an important question about the underlying mechanisms of crossmodal binding. We have shown that crossmodal interactions gracefully degrade as the discrepancy between the two sensory signals increases [10]. This behavior is predicted by Bayesian causal inference model reviewed in the target article, as the larger the discrepancy between the two signals the lower the probability that they were caused by the same object. As noted by Spence [2], the binding of two signals also depends on the prior expectation of a common cause. These are two factors that determine the probability of sensory signals having a common cause [10,13,17,18]. Spence [2] identifies attention as a likely important factor in crossmodal feature binding. Indeed some studies suggest that attention must be spread across modalities for multisensory integration to occur; selectively attending to one modality may attenuate integration effects and possibly inhibit processing of the unattended modality. Such effects on multisensory interactions have been reported behaviorally [19–21] and in an ERP study [22]. On the other hand, the findings of multisensory effects in the absence of awareness of extramodal stimuli [23], and enhancement of visual cortex excitability before conscious discrimination of sounds [24], discussed by de Haas and Rees in their commentary, suggest that attention may not be needed for certain multisensory effects. The role of attention in multisensory binding and integration is still an important open question that merits further examination.

References

[1] de Haas B, Rees G. Multiple stages of cross-modal integration in visual processing: Comment on "Crossmodal influences on visual perception" by L. Shams & R. Kim. Phys Life Rev 2010;7(3):287–8, this issue.

[2] Spence C. Multisensory integration – Solving the crossmodal binding problem: Comment on "Crossmodal influences on visual perception" by Shams & Kim. Phys Life Rev 2010;7(3):285–6, this issue.

[3] Alais D. Helping the visual system find its target: Comment on "Crossmodal influences on visual perception" by L. Shams & R. Kim. Phys Life Rev 2010;7(3):293–4, this issue.

[4] Vroomen J. Causal inference in audiovisual speech: Comment on "Crossmodal influences on visual perception" by L. Shams. Phys Life Rev 2010;7(3):289–90, this issue.

[5] Barone P. Is the primary visual cortex multisensory? Phys Life Rev 2010;7(3):291–2, this issue.

[6] Shams L, Kim R. Crossmodal influences on visual perception. Phys Life Rev 2010;7(3):269–84, this issue.

[7] Watkins S, Shams L, Josephs O, Rees G. Activity in human V1 follows multisensory perception. NeuroImage 2007;37:572–8.

[8] Watkins S, Shams L, Tanaka S, Haynes J-D, Rees G. Sound alters activity in human V1 in association with illusory visual perception. NeuroImage 2006;31:1247–56.

[9] Shams L, Iwaki S, Chawla A, Bhattacharya J. Early modulation of visual cortex by sound: An MEG study. Neurosci Lett 2005;378(2):76–81.

[10] Körding K, Beierholm U, Ma WJ, Tenenbaum JM, Quartz S, Shams L. Causal inference in multisensory perception. PLoS ONE 2007;2:e943.

[11] Helbig HB, Ernst MO. Knowledge about a common source can promote visual-haptic integration. Perception 2007;36:1523–33.

[12] Shams L, Beierholm U. Causal inference in perception. Trends Cogn Sci [in press].

[13] Shams L, Ma WJ, Beierholm U. Sound-induced flash illusion as an optimal percept. Neuroreport 2005;16:1923–7.

[14] Wozny D, Beierholm U, Shams L. Probability matching as a computational strategy used in perception. PLoS Comput Biol 2010;6(8):e1000871. doi:10.1371/journal.pcbi.1000871.

[15] Fishman MC, Michael P. Integration of auditory information in the cat's visual cortex. Vision Res 1973;13(8):1415–9.

[16] Morrell F. Visual system's view of acoustic space. Nature 1972;238(5358):44–6.

[17] Beierholm U, Quartz S, Shams L. Bayesian priors are encoded independently of likelihoods in human multisensory perception. J Vis 2009;9(5):1–9.

[18] Wozny DR, Beierholm UR, Shams L. Human trimodal perception follows optimal statistical inference. J Vis 2008;8(3):1–11.

[19] Mozolic JL, Hugenschmidt CE, Peiffer AM, Laurienti PJ. Modality-specific selective attention attenuates multisensory integration. Exp Brain Res 2008;184(1):39–52.

[20] Oruc I, Sinnett S, Bischof WF, Soto-Faraco S, Lock K, Kingstone A. The effect of attention on the illusory capture of motion in bimodal stimuli. Brain Res 2008;1242:200–8.

[21] Sinnett S, Costa A, Soto-Faraco S. Manipulating inattentional blindness within and across sensory modalities. Q J Exp Psychol (Colchester) 2006;59(8):1425–42.

[22] Talsma D, Doty TJ, Woldorff MG. Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? Cereb Cortex 2007;17(3):679–90.

[23] Dufour A, Touzalin P, Moessinger M, Brochard R, Despres O. Visual motion disambiguation by a subliminal sound. Conscious Cogn 2008;17(3):790–7.

[24] Romei V, Murray MM, Cappe C, Thut G. Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. Curr Biol 2009;19(21):1799–805.