
http://www.diva-portal.org

Postprint

This is the accepted version of a paper presented at *Cognitive Vision: Integrated Vision and AI for Embodied Perception and Interaction. Workshop on Cognitive Vision at Advances in Cognitive Systems (ACS), MIT, Boston, USA, August 2, 2019*.

Citation for the original published paper:

Hemeren, P. (2019)
To AIR is Human, or is it?: The Role of High-Level Representations and Conscious Awareness in Biological Motion Perception
In:

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:his:diva-18347

To AIR is human, or is it?
The role of high-level representations and conscious awareness in biological motion perception

Paul E. Hemeren (paul.hemeren@his.se)

University of Skövde, School of Informatics
Box 408, 541 28, Skövde Sweden

The purpose of this research is to address the nature of high-level processing within visual perception. In particular, results from the visual processing of biological motion will be used to discuss the role of attention in high-level vision and visual consciousness. Original results from 3 priming experiments indicate "automatic" high-level semantic activation in biological motion perception. The view presented here is discussed in the context of Prinz's (2000, 2003) AIR-theory. AIR stands for Attended Intermediate-level Representations and claims that visual consciousness resides at the level of intermediate-level representations. In contrast, the view presented here is that results from behavioral and neuroscientific studies of biological motion suggest that visual consciousness occurs at high cortical levels. Moreover, the Reverse Hierarchy Theory of Hochstein and Ahissar (2002) asserts that spread attention in high cortical areas is indicative of what they term "vision at a glance." The gist of their theory is that explicit high-level visual processing involves initial feedforward mechanisms that implicitly follow a bottom-up hierarchical pathway. The end product of the processing, and the beginning of explicit visual perception, is conscious access to perceptual content in high-level cortical areas. Finally, I discuss the specific claims in AIR and present objections to Prinz's arguments for why high-level visual processors are not good candidates for the locale of consciousness. In conclusion, the central claim of AIR with an emphasis on the connection between intermediate level representations and perceptual awareness seems to be too strong, and the arguments against high-level perceptual awareness are not convincing.

Visual processing of biological motion

It's no small secret that human vision is highly sensitive to the motion patterns created by the movement of other individuals. This sensitivity, however, is not restricted to motion patterns as such. When we see the movements of others, we do not merely see the independent movement of hands, arms, feet and legs and movement of the torso. Instead, we are able to quickly and accurately identify motion patterns as meaningful actions (running, jumping, throwing, crawling, etc.). Gunnar Johansson (1973, 1975) clearly demonstrated an effective method for investigating the sensitivity of human action perception using the point-light technique. By placing small lights or reflective patches on the joints of a human actor dressed in dark clothing and filming various actions, Johansson could isolate the motion (kinematic) information associated with the different actions by adjusting the contrast of the filmed sequences so that only the points of light were visible to human observers. When just one (static) frame from one of these motion sequences was presented to observers, they were unable to discern any meaningful representation. When, however, consecutive frames were displayed to produce (apparent) motion, observers could immediately see the portrayed action, walking or a couple dancing a Swedish folk dance. Johansson (1975) states, "that as little as a tenth of a second (the time needed to project two motion-picture frames) is often enough to enable a naïve observer to identify a familiar biological motion."

An important feature of the visual processing of the dynamic human gestalt in point light displays is the "automatic" nature of the perceptions. As Johansson (1973) points out, "… we have found that it seems to be a highly mechanical, automatic type of visual data treatment that is most important." While Johansson's use of the term "automatic" points more to the early processes involved in establishing hierarchies of locally rigid perceptual units, there is a case to be made for the automatic processing of biological motion at a higher cognitive level. Phenomenally, Johansson's own demonstrations point to the immediateness and vividness of viewing point-light displays of biological motion. Observers are fast and accurate in their identifications. They appear to have direct access to a level of meaning or semantic level representation of the actions depicted in the point-light displays. Runeson and Frykholm (1983) demonstrated that participants are sensitive to the dynamics (not direct visible) that constrain the motion patterns in the point-light displays. Participants in their studies could reliably detect the various weights of an unseen box lifted by the point-light actor. The length of a throw could also be determined on the basis of the pattern of motion of the thrower. Even the gender of adults and children were detected with reliable accuracy. So it seems humans have direct/automatic phenomenal access to action meaning.

The assertion that biological motion perception is fast, categorical and automatic is supported by a number of studies. Jokisch, Daum, Suchan and Troje (2005) measured event-related potentials when subjects viewed a point-light walker. Their results showed that the ability to discriminate between upright and inverted displays occurred within 180 ms. Giese and Poggio (2003) have developed and implemented a computational model based on behavioral and neuroscientific data for visual processing of biological motion. The model, which builds upon the classification of neuronal populations from the dorsal and ventral streams, accurately responds to different types of actions, exhibiting categorization. A central assumption in their model is the feedforward hierarchical processing indicative of the dorsal and ventral streams. The simulations also show fast processing (< 200 ms) of the stimuli where activity increases quickly after stimulus onset. This result is consistent with results from electrophysiological experiments showing selective activity within less than 200 ms for some neurons in the superior temporal sulcus (STS), a high-level cortical area.

The automatic nature of biological motion processing has been shown in priming studies that I have conducted. Subjects were presented with upright and inverted point-light displays of three different actions. The task was to simply indicate whether the display was upright or inverted. An assumption of this orientation decision task was that it did not require semantic level processing. Given this assumption, if priming effects show that different actions prime one another differently, then it would appear that subjects are indeed using semantic level information in their decisions that allow them to discriminate between the different kinds of actions. The results showed that this is what happened. The crucial finding in this experiment showed the same level of priming for inverted displays *but* there was significantly less indication that subjects were discriminating between the different kinds of actions. So, for upright displays there is an indication that subjects have automatic access to semantic level information for upright displays but not for inverted displays.

Given the fast and automatic nature of biological motion perception, it may seem that attention is not needed. This, however, is not the case. Studies by Thornton, Rensink and Shiffrar (2002), Cavanagh, Labianca and Thornton (2001) and Battelli, Cavanagh and Thornton (2003) all demonstrate results indicating that attention is necessary. For example, Battelli et al. tested the perception of biological motion on 3 parietal patients to assess the role of attention and whether or not visual processing of biological motion requires the activation

of higher cortical areas extended beyond (low-level) motion processing areas. The human homologue to MT/V5 (motion processing) was apparently spared to some extent in these patients as indicated by their normal performance on a low-level motion task. In contrast, patients with damage to MT/V5 fail in similar tasks. In a visual search task using point-light walkers, the patients were severely impaired as revealed by decreasing performance according to distractor set size. Even when presented with a single point-light figure, the patients had difficulty in detecting the direction of the point-light walker. The fact that the subjects were parietal patients suggests attentional deficits. And because they performed poorly on the point-light stimuli, this implicates a role of attention in biological motion perception. The results also indicate a higher level of processing is necessary than the extrastriate MT/V5 region.

What happens to biological motion processing when the above mentioned extrastriate region is damaged? Mcleod, Dittrich, Driver, Perett and Zihl (1996), Vaina, Lemay, Bienfang, Choi and Nakayama (1990) investigated this question in two patients with lesions to MT/V5. While severely impaired on low-level motion tasks, the patients demonstrated a remarkable ability to process biological motion. In the Mclead et al. study, the patient was essentially motion blind. In a further study (Vaina & Gross, 2004) on four patients with temporal lobe damage that likely included the human homologue of STP (superior temporal polysensory area) the results showed severe impairments on biological motion for all patients. It should, however, be mentioned that the results from psychophysical motion tasks showed differing results. So some low-level motion processing was intact for some of the patients.

The conclusion to be reached on the basis of the reported studies is that there is reason to believe that higher cortical areas play a central role in biological motion processing. It also appears to be the case that attention is necessary for this processing. Given these findings, it seems reasonable to conclude that phenomenal awareness also occurs at a high-level for biological motion perception. But according to Prinz (2000, 2003) phenomenal awareness occurs *only* at an intermediate-level of visual processing. This idea is formulated in his AIR theory of consciousness.

The AIR theory

The AIR theory takes Marr's (1982) theory of vision as a starting point. According to Prinz, Marr's intermediate-level (2½D sketch) is where information about surfaces, depth and shape are encoded. A further property is the viewer-centered representations on that level of visual processing. Prinz presents arguments for why he thinks this intermediate-level is the locale of visual consciousness. He discusses why both low- and high-level representations are not good candidates for visual consciousness. But before presenting his specific arguments, I will briefly state his main claim. "*Conscious [sic] seems to arise in intermediate-level perceptual subsystems when and only when activity in those systems is modulated by attention. When attention is allocated, perception becomes conscious. Attentional modulation of intermediate level representations is both necessary and sufficient for consciousness*." (Prinz, 2003, p. 4) Hence AIRs (Attended Intermediate-level Representations).

My claim is that biological motion perception represents a perceptual phenomenon that appears to be an exception to AIR. Therefore AIR is too strong in its claims. For the sake of clarity, my argument is presented in the following points:

1. Results from biological motion processing for normal upright displays show that categorical discriminations are fast (Jokish et al., 2005; Giese & Poggio, 2003) and "automatic" (Giese & Poggio, 2003) as revealed in short-term priming (Hemeren).

2. Biological motion perception requires attention (Thornton et al., 2002; Cavanagh et al., 2001; Battelli et al., 2003).

3. Attention is necessary and sufficient for visual consciousness (Prinz, 2003).

4. Biological motion perception is high-level (implicated in processing levels beyond extrastriate motion areas) (Battelli et al., 2003; Mcleod et al., 1996; Vaina et al., 1990; Jokish et al., 2005).

Therefore, it is not the case that "consciousness seems to arise in intermediate-level perceptual subsystems when and only when activity in those systems is modulated by attention" (Prinz, 2003).

Despite appearances, this argument still has at least one hole (there are probably others), and Prinz, being a trained philosopher, will likely find it. In a strict sense, he can still accept the premises without necessarily being committed to the conclusion. He could for example say, "Yes, biological motion processing does have a high-level component, and it does require attention, and there is phenomenal awareness, but this does not rule out the possibility that there also are intermediate-level perceptual subsystems being modulated by attention, and these, not high-level areas, are doing the awareness work." Being on my guard, I'm ready for this objection. First, I do not want to claim that no intermediate-level perceptual subsystems in biological motion processing are modulated by attention. It is indeed quite likely that feedback mechanisms directed by attention modulate perception at an intermediate level. The problem with this objection is that I know of no one who asserts (in the case of biological motion perception) that it is when and only when subsystems on this level are modulated by attention that phenomenal awareness arises. My reply here is a burden-of-proof-reply, and as such I'm open to evidence that *clearly* supports that claim.

Let me also be clear about another claim that I am not making. I am not claiming that there are no intermediate-level perceptual subsystems of the variety that Prinz is proposing. Perhaps color vision is one such domain, and object recognition may be another, although I have my doubts. The upshot is that Prinz's central claim in AIR is too strong given current results from the domain of biological motion perception.

In addition to biological motion perception as a counter-example to AIR, there are more internal criticisms to be raised. Why does Prinz think that high-level representations are not a plausible candidate for the locale of visual consciousness? He gives us 3 main reasons:

1. Since high-level vision is object-centered (Marr, 1982) and our visual experience is viewer centered, it cannot be the case that high-level representations are the locale of visual consciousness.

   Objection: There is still no common agreement regarding the extent to which high-level vision is object centered, there are actually cells that are viewer-centered and others that are object centered in high-level processing.

   Objection: Why can't top-down mechanisms arising in higher levels modulate view sensitive cells in lower visual areas? Hochstein and Ahissar (2002) assert that this is quite likely what happens, although they do not discuss object perception per se.

2. Associative agnosia is used as evidence that we can still "have visual phenomenology without access to high-level representations in IT." The example he gives is that despite the lack of an ability to recognize objects, patients can accurately draw pictures of objects.

Objection: If this is the case, it would seem that extrastriate areas are intact, and the question then becomes "What phenomenal experience is then associated with accurately drawing the pictures?" What is needed to support the strong claim in AIR is a double dissociation between high and intermediate level processing/representation. Specifically, the requirement is to show that when damage occurs in high cortical levels and intermediate levels are intact that there is perceptual awareness *and* the converse, i.e., intact high levels and damaged intermediate levels and no perceptual awareness. If the theory holds then this should be the case for all perceptual processing accompanied by awareness. Associative agnosia may be an example of just one side of the double dissociation if Prinz is correct. He still needs to provide evidence for the other dissociation and give us some good reasons why this should be the case for all forms of perceptual awareness.

3. In his 2003 paper, Prinz suggests a further reason why the high-level is not a good candidate. "Cells in IT encode the wrong information. They tend to abstract away from size and specific orientation."

Objection: Cells in IT (inferior temporal area, high-level) certainly would be encoding the wrong information *if* one assumes that the intermediate level is the locale of consciousness and perception on that level is viewer-centered. But that seems to be begging the question. AIR theory is a theory about where visual consciousness is located. To assume the theory itself in order to make the claim about IT-cells encoding the wrong information and then using that as a support for the theory seems dubious.

The arguments against high-level processing mechanisms playing a central role in visual consciousness are not convincing. It is tenuous ground on which to discount the phenomenal role of high-level representations. I'm inclined to think that there are other factors that need to be taken into account when investigating the relationship between visual consciousness, perception and neuroscience. These other factors have to do with the behavioral and ecological validity of the theoretical claims being made. And my point is that explicit high-level perception is where the proverbial 'rubber meets the road' in human visual cognition. The fact that changes in low-level and intermediate level neuronal activity are modulated by an animal's behavioral adaptation to stimulus contingencies (Lee, Yang, Romero & Mumford, 2002) suggests functional visual awareness on a high-level of processing, including high-level representations that importantly figure into ecologically guided potential for action. And the perception and representation of actions are prime candidates for exemplifying the intricate link between perception and ecological factors broadly defined. Results from neuroscience and behavioral studies support this view. People are keenly aware of the physical and dynamic constraints that govern how the human body moves. Dynamic factors that influence the kinematic patterns in the point-light displays are readily picked up by human observers (Runeson & Frykholm, 1983). Research on representational momentum also shows that our representations of events and human movement include dynamic factors. (See for example Kourtzi & Shiffrar, 1999).

If I were to really stretch these observations into a more speculative view of human cognition, I would be tempted to view issues of visual awareness in the service of action, perhaps a broadly defined definition of action. In this sense, cognition is not only grounded in perception but is grounded in the role that perception plays in an organism's ability to interact with its social and physical environment. It is not simply the case that we act to perceive but rather that we perceive to act. And to the extent that this conjecture holds, human cognition

and it's contextually guided ecological setting is where to begin to look for a locale of visual consciousness or phenomenal awareness (Clark, 1999).

References

Battelli, L., Cavanagh, P. & Thornton, I. (2003). Perception of biological motion in parietal patients. *Neuropsychologia*, 41, 18808-1816.

Cavanagh, P., Labianca, A.T., & Thornton, I. (2001). Attention-based visual routines: Sprites. *Cognition*, 80, 47-60.

Clark, A. (1999). Awareness and visuomotor action. *Journal of Consciousness Studies*, 6, 1-18.

Giese, M. & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nat Rev Neuroscience*, 4(3), 179-192.

Hochstein, S. & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36, 791-804.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14, 201-211

Jokisch, D., Daum, I., Suchan, B. & Troje, N. F. (2005). Structural encoding and recognition of biological motion: evidence from event-related potentials and source analysis. *Behavioural Brain Research*, 157, 195-204.

Kourtzi, Z. & Shiffrar, M. (1999).Dynamic representations of human body movement. *Perception*, 28, 49-62.

Lee, T.S., Yang, C.F., Romero, R.D. & Mumford, D. (2002). *Nature Neuroscience*, 5(6), 589-597.

Marr, D. (1982). *Vision*. New York: W.H. Freeman.

Mcleod, P., Dittrich, W., Driver, J., Perrett, D. & Zihl, J. (1996). Preserved and impaired detection of structure from motion by a "motion-blind" patient. *Visual Cognition*, 3(4), 363-391.

Prinz, J. (2000). A neurofunctional theory of visual consciousness. *Consciousness and Cognition*, 9, 243-259.

Prinz, J. (2003). A neurofunctional theory of consciousness. In A. Brook and K. Akins (Eds.) *Philosophy and neuroscience*, Cambridge: Cambridge University Press.

Runeson, S. & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender, recognition, and deceptive intention. *Journal of Experimental Psychology: General*, 112(4), 585-615.

Thornton, I. M., Rensink, R. A. & Shiffrar, M. (2002). Active and passive processing of biological motion. *Perception*, 31, 837-853.

Vaina, L.M., Lemay, M., Bienfang, D.C., Choi, A.Y. & Nakayama, K. (1990). Intact "biological motion" and "structure from motion" perception in a patient with impaired motion mechanisms: A case study. *Visual Neuroscience*, 5, 353-369.

Vaina, L.M. & Gross, C.G. (2004). Perceptual deficits in patients with imparied recognition of biological motion after temporal lobe lesions. *Proc. Natl. Acad. Sci. USA*, 101(48), 16947-16951.