Reverse Hierarchy Theory and the Role of Kinematic Information in Semantic Level Processing and Intention Perception[1]

Paul Hemeren

University of Skövde, Skövde, Sweden

In many ways, human cognition is importantly predictive (e.g., Clark, 2015). A critical source of information that humans use to anticipate the future actions of other humans and to perceive intentions is bodily movement (e.g., Ansuini et al., 2014; Becchio et al., 2018; Koul et al., 2019; Sciutti et al., 2015). This ability extends to perceiving the intentions of other humans based on past and current actions. The purpose of this abstract is to address the issue of anticipation according to levels of processing in visual perception and experimental results that demonstrate high-level semantic processing in the visual perception of various biological motion displays. These research results (Hemeren & Thill, 2011; Hemeren et al., 2018; Hemeren et al., 2016) show that social aspects and future movement patterns can be predicted from fairly simple kinematic patterns in biological motion sequences, which demonstrates the different environmental (gravity and perspective) and bodily constraints that contribute to understanding our social and movement-based interactions with others. Understanding how humans perceive anticipation and intention amongst one another should help us create artificial systems that also can perceive human anticipation and intention.

A key theoretical basis in this research is the Reverse Hierarchy Theory (Ahissar & Hochstein, 2004), which proposes that perceptual learning begins at high perceptual levels (vision-at-a-glance) as a result of (relatively) fast implicit feedforward processing. This processing results in conscious perception. With greater exposure and expertise,

[1] This work has been financially supported by the Knowledge Foundation, Stockholm, under SIDUS grant agreement no. 20140220 (AIR, \Action and intention recognition in human interaction with autonomous systems").

access to lower-level information can be triggered when vision-at-a-glance fails. The more difficult a perceptual task becomes, the greater the probability of the need to access lower-level information in order to solve the perceptual task, which is termed vision-with-scrutiny (Ahissar & Hochstein, 2004). A critical feature of this theory is that high-level representations "facilitate the identification of ecologically relevant elements," (p.463). Such ecologically relevant elements can be complex features consistent with sematic level processing, which includes anticipation and intention as well as initial conscious perception.

In a series of experiments, I and my colleagues have used psychophysical methods and recordings from interactions with objects in natural settings. This includes experiments on the incidental processing of biological motion (Veto et al., 2013), social gesture recognition (Hemeren et al., 2016) as well as studies that examine the role of kinematic patterns of cyclists and driver's accuracy to predict the cyclist's intentions in traffic (Hemeren et al., 2014).

The results show both clear effects of "low-level" biological motion factors, such as opponent motion, on the incidental triggering of attention in basic perceptual tasks and "higher-level" top-down perception in the intention prediction of cyclist behavior.

Within the context of the program theme, the above results will be used to indicate the interplay between expectation mediated (top-down) and stimulus driven effects of visual processing in the context of human interaction.

The discussion will include the role of context in gesture recognition and the extent to which machine learning techniques can use kinematic information to perceive human intentions. This research builds on our previous results that demonstrated a strong association between the top-down activation of an action representation and the kinematics of the specific grasping action. Experimental results on the classification of gesture stimuli by humans and by four machine-learning techniques demonstrate the critical relationship between action kinematics and judgments of grasping and the social quality of the hand/arm gestures. Our results support previous research on intention-from-movement understanding (e.g., Ansuini et al., 2014; Becchio et al., 2018; Koul et al., 2019; Sciutti et al., 2015) that demonstrates the reliance on kinematic information for perceiving the intentions in different grasping actions as well as communicative point-light actions.

By using point-light displays of biological motion (e.g., Johansson, 1973) in experiments on the visual perception of human actions, the systematic influence of visible kinematic local and global motion can be studied. The original findings from Johansson (1973) demonstrate that when presented with a static form, people have difficulty in identifying the figure and action. However, once the figure starts to move, people see the action that the person is performing. Much previous research (see e.g., Shiffrar & Pinto, 2002) demonstrates the holistic/global processing involved in the visual perception of point-light displays of biological motion. From a perceptual learning perspective, biological motion perception is an example of the Eureka effect (Ahissar & Hochstein, 2004), in which learning is governed by top-down control and single exposures and has long-lasting effects.

Experiments on action perception in point-light displays of biological motion can be used as a basis to develop models for predicting the intentions of cyclists in traffic (Hemeren et al., 2014). Different motion parameters are likely differently critical for different actions, and the perceptual saliency of different patterns of bodily movement will signal different intentions. In this research, we addressed the following questions: 1) How accurate are human observers at predicting the behavior of cyclists as the cyclists approached a crossing? 2) If the accuracy is reliably better than chance, what cues were used to make the predictions? 3) At what distance from the crossing did the most critical cues occur? 4) Can the cues be used in a model that can reliably predict cyclist intent? In these experiments, people observed the behavior of cyclists approaching a crossing and then made predictions about whether or not the cyclist would turn or go straight on.

The results from the human observers showed a clear reliance on a few critical movement parameters for predicting cyclist intent. Head turning and pedaling for example seem to be critical signals for predicting turning behavior whereas speed and speed change are more critical for predicting that a cyclist will continue straight. It is these context-based parameters that allow drivers to anticipate cyclist behavior. Drivers also appear to detect the biological motion of cyclists even when other competing sources of information also are visible. The results point to the critical role that biological motion can play on predicting the intention and detection of cyclists in traffic. This information can be used to inform (semi-)autonomous systems of human intentions in traffic.

In summary, the contribution of these research results is that human kinematic motion can be processed implicitly using fast feedforward connections that lead to a high-level semantic understanding of human movement, which includes categorizing actions and perceiving social communication. It is only when a central task becomes difficult that more attention is needed to search for more information using explicit feedback connections according to Reverse Hierarchy Theory (Ahissar & Hochstein, 2004). Development in AI for creating reliable intelligent systems for predicting human behavior and intentions will therefore have to at least appear to focus on achieving this high-level understanding of human motion in relation to past and current environments. It is certainly logically possible that even given the knowledge we have about how humans perceive other humans that AI technology may be able to achieve a similar behavior without access to any fast implicit feedforward processing that leads to high-level semantics.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci*, Oct; 8(10): 457-64.

Ansuini, C., Cavallo, A., Bertone, C., & Becchio, C. (2014). The visible face of intention: why kinematics matters. *Frontiers in Psychology*, 5:815. doi: 10.3389/fpsyg.2014.00815

Becchio, C., Koul, A., Ansuini, C., Bertone, C., & Cavallo, A. (2018). Seeing mental states: An experimental strategy for measuring the observability of other minds. *Physics of life reviews*, 24, 67-80.

Clark, A. (2015). Embodied Prediction. In T. Metzinger & J.M. Windt (Eds). *Open MIND*: 7(T). doi: 10.15502/9783958570115

Hemeren, P., Johannesson, M., Lebram, M., & Eriksson, F. (2018). Signals for Active Safety Systems to Detect Cyclists and Their Intentions in Traffic. Platform talk at the 8th World Research Congress, The Eye, The Brain, and The Auto. Detroit, Michigan.

Hemeren, P., Johannesson, M., Lebram, M., Eriksson, F., Veto, P., & Ekman, K. (2014). The Use of Visual Cues to Determine the Intent of Cyclists in Traffic. *IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support* (CogSIMA), March 3-6, San Antonio, USA.

Hemeren, P.E., & Thill, S. (2011). Deriving motor primitives through action segmentation. *Frontiers in Psychology*, 1:243. doi: 10.3389/fpsyg.2010.00243

Hemeren, P., Thill, S., Lee, G. & Veto, P., (2016). Kinematic specifications of grasping and social dimensions in hand actions. Workshop on Communicating Intentions in Human-Robot Interaction in conjunction with the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2016), Columbia University, New York.

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

Koul, A., Soriano, M., Tversky, B., Becchio, C., & Cavallo, A. (2019). The kinematics that you do not expect: Integrating prior information and kinematics to understand intentions. *Cognition*, 182, 213-219.

Sciutti, A., Ansuini, C., Becchio, C., & Sandini, G. (2015). Investigating the ability to read others' intentions using humanoid robots. *Frontiers in Psychology*, 6:1362. doi:10.3389/fpsyg.2015.01362

Shiffrar, M. & Pinto, J. (2002) The visual analysis of bodily motion. In W. Prinz & B. Hommel (Eds.) *Common mechanisms in perception and action: Attention and Performance*. Oxford: Oxford University Press.

Veto, P., Thill, S., & Hemeren, P. (2013). Incidental and non-incidental processing of biological motion: Orientation, attention and life detection. In M. Knauf, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.) *Proceedings of the 35th Annual Conference of the Cognitive Science Society*, 1528-1533. Austin TX: Cognitive Science Society.