position of our moving hand with the position of the moving ball. However, an opposite position that is not often entertained is that the goal of the visual system is to deliver to the animal visual information that is suitably shaped for action directed at a moving ball. Thus, during a catch, the visual system modifies the perceived position of the ball so that it matches the position of the moving hand.[3] On this view, the goal of the visual system is to generate an output that has shared features with motor processes: in particular, predictions (Wilson & Knoblich 2005).

The proposal is that visual representations that receive feedback from higher cortical areas are susceptible to modification. Thus, these visual representations are controlled entities, just like neural limb representations. The descending visual signals cannot, of course, activate otherwise silent neurons, which is presumably only possible on the basis of stimulus-driven retinal input (Hupe et al. 1998). But the descending signals can, nonetheless, affect ongoing activity in many areas (e.g., the thalamus) and produce anticipatory spatial adjustments (Sillito et al. 1994).

Although my somewhat limited goal was to evaluate the feasibility of visual prediction, during the course of this endeavor it seems that prediction may be far more pervasive in the CNS than originally expected. The novel approach developed here may be easily adapted to investigate predictive phenomena in the CNS more generally. Visual prediction has a strong logical basis and seems consonant with other visual phenomena such as the various constancies and motion deblurring, as well as theoretical constructs such as neural competition. Prediction may be a multi-level, multi-modal phenomenon found in both sensory and motor systems. Furthermore, prediction may result from computations carried out by single neurons, or neural networks, or both. This general approach to the study of prediction suggests possibilities that could unify research from single cells to cognition.

NOTES
1. Since signals within the retina itself are transmitted via graded potentials, as opposed to action potentials, the time taken for lateral interactions across 0.15 mm of retina could take significantly shorter time than the above estimate (Bruce Bridgeman 2005, personal communication).
2. There could be many reasons for behaviors normally displayed by animals to break down. Breakdown in behavior is known to occur, for example, in nervous systems in which disease has affected neural processing delays. For example, Multiple Sclerosis occurs because of demyelination, which affects neural transmission delays. Commonly observed behavior in healthy animals leads to the following assumption: In the absence of mechanisms compensating for neural delays, many behaviors in otherwise healthy animals would be disrupted. An analogous assumption holds for visual delays. The fundamental assumption of compensation for visual delays states that in the absence of mechanisms compensating for visual delays, many behaviors in otherwise healthy animals would be disrupted. Note

that this last statement makes no assumption about whether visual or non-visual mechanisms compensate for visual delays.
3. One significant difference between the treatments of the visual position of a moving object and the sensed position of a moving limb is that we appear to have no conscious control over the position of the moving object, while we do have conscious control over the position of our limb. However, we are aware of only some internal representations that allow us to predict the future states of our limbs from current states during movement (Blakemore et al. 2002). Thus, many representations that allow for prediction of moving visual objects and of limbs during movement are not available to awareness.

Open Peer Commentary

Shifting attention to the flash-lag effect

Marcus Vinícius C. Baldo[a] and Stanley A. Klein[b]
[a]*"Roberto Vieira" Laboratory of Sensory Physiology, Department of Physiology and Biophysics, Institute of Biomedical Sciences, University of São Paulo, São Paulo, SP 05508-900, Brazil;* [b]*Visual Processing Laboratory, School of Optometry, University of California at Berkeley, Berkeley, CA 94720-2020.*
baldo@usp.br   sklein@berkeley.edu
http://www.fisio.icb.usp.br/~vinicius/index.html

Abstract: An attention shift from a stationary to a changing object has to occur in feature – rather than physical – space, in order to bind these stimuli into a unitary percept. This time-consuming shift leads to the perception of a changing stimulus further ahead along its trajectory. This attentional framework is able to accommodate the flash-lag effect in its multiple empirical manifestations.

The flash-lag effect (FLE) is an empirical fact. Understanding the effect, however, requires us to build conceptual models that can range from the biophysical description of neuronal interactions to global cognitive schemes. Whichever model may be conceived, it should explain the FLE and related findings, rely on sound physiological grounds, adopt the smallest set of assumptions, and predict novel phenomena.

Motion extrapolation, as originally proposed by Romi Nijhawan (1994), nicely explains the flash-lag phenomenon in its purest form. However, in order to accommodate that explanation with discordant empirical findings, Nijhawan, in addition to splitting perceptual extrapolation in two successive steps, now resorts to at least three different putative mechanisms: (1) the role of lateral communications between neighboring retinotopic locations, (2) the impact of visual transients generated by offsets, and (3) a dynamical interplay between motor and visual processes. Therefore, Nijhawan's effort to fit motion extrapolation into challenging empirical data has compelled him to adopt a piecewise mosaic of physiological functions. Also, even though Nijhawan does a good job in weaving the explanatory content of his account, he leaves aside a deeper exploration of its predictive potential.

Both empirical and theoretical evidences point to the ability of basic sensory operations to generate the FLE (Baldo & Caticha 2005; Berry et al. 1999; Erlhagen 2003). In fact, a simple neural network model built upon a small set of physiologically grounded assumptions was able to replicate the standard FLE (under several conditions), as well as its dependence on stimuli luminance, priming, trajectory, and predictability (Baldo & Caticha 2005). This model also helps us evaluate the relative

underlying involvement of spatial sources (lateral interactions between neighboring receptive fields) and temporal sources (time delays required to activate the neurons in the network or to broadcast signals among them).

However, despite their physiological appeal, simplistic models are still unable to encompass the phenomenological richness portrayed by this illusion, such as has been observed in the chromatic (Nijhawan 1997; Sheth et al. 2000), spatial frequency (Sheth et al. 2000), and auditory (Alais & Burr 2003) domains. This wide-ranging manifestation of the FLE would call for the contribution of higher, more integrative, underlying mechanisms. Over the last few years, we have been refining a conceptual framework in which attention plays a role in contributing to the FLE. As pointed out by Nijhawan himself, "distances that separate the representations are large, so the information exchange between the representations is time consuming" (target article, sect. 9.1, para. 1). Actually, the first conceptual alternative challenging the motion extrapolation account was based on the time consumed by a unidirectional shift of (spatial) attention from the flashed location to the moving location (Baldo & Klein 1995). Meanwhile, whereas the role of attention in causing the FLE remained elusive, the modulation of the FLE by attentional factors was gradually being established (Baldo & Namba 2002; Baldo et al. 2002; Chappell et al. 2006; Namba & Baldo 2004; Sarich et al. 2007).

More recently, we have advanced our attentional framework by proposing that the FLE could originate from the time needed for attention to bind the flash and moving percept into a unitary whole (Baldo & Klein, in press). Starting with the detection of an abrupt event (a stationary flash or a beep, for instance), a shift of attention from a stationary object to a changing ("moving") object has to occur in feature rather than physical space in order to bind them together into a unitary percept. This object-based attentional shift would require some time to be carried out, regardless of any spatial separation between both visual stimuli, and would naturally lead to a percept consisting of a changing stimulus further ahead along its "trajectory" (equivalently, we can also consider a non-directional attentional spread over the object, in feature space, which starts when an abrupt-onset stimulus is presented – though not necessarily from where it is presented).

It is easy to see that this feature-based "attentional" explanation is not in conflict with the findings concerning the flash-initiated and flash-terminated conditions, as claimed before in relation to the purely spatial attention shift (Alais & Burr 2003; Khurana & Nijhawan 1995): Whereas the time required for the attentional binding will lead to the FLE in the former condition, no FLE will be observed in the latter (in opposition to motion extrapolation's predictions), since the moving stimulus never reaches a position beyond that where it disappeared or stopped.

The relationship between delays required to bind a spatially extended object and time-consuming shifts (or spread) of attention has been extensively reported (Chappell et al. 2006; Enns & Oriet 2004; Houtkamp et al. 2003; Kanai et al. 2004; Roelfsema et al. 2000; Sarich et al. 2007). Because the scheme we propose involves an attentional binding in feature space, our account unifies empirical findings observed in a broad class of visual features (such as position, color, luminance, and spatial frequency), as well as in crossmodal phenomena (Alais & Burr 2003). It is worth noting that only cognitive-oriented models are presently able to capture the entire perceptual spectrum revealed by the FLE. Nijhawan's motion extrapolation and the attentional binding account we presently offer are the main conceptual frameworks embodying this class of integrative models. Our proposal, however, is favored by a more parsimonious set of assumptions, a deeper physiological root, and a greater predictive power.

Eventually, cognitive models are to be brought down to the wiring structure and functional operations of neuronal lattices.

It is conceivable that lateral interactions giving rise to spatio-temporal facilitations, as evidenced by simple networks (Baldo & Caticha 2005) and also assumed by Nijhawan, could be further elaborated and extended to perceptual processing other than visual. A desirable connection between the fundamental rules of neural functioning and the large-scale expression of cognitive integration could thus be envisaged.

In conclusion, it is encouraging to recognize that the decade-long dispute over the underpinnings of the FLE has already contributed toward refining, discarding, or amalgamating fragmented and contradictory conceptions. Whatever the final consensus turns out to be, Nijhawan's legacy to the current debate on visual perception will endure.

Anticipation requires adaptation

Christian Balkenius and Peter Gärdenfors
Department of Cognitive Science, Lund University, Kungshuset, Lundagård, 222 22 Lund, Sweden.
christian.balkenius@lucs.lu.se   **http://www.lucs.lu.se**
peter.gardenfors@lucs.lu.se   **http://www.lucs.lu.se**

Abstract: To successfully interact with a dynamic world, our actions must be guided by a continuously changing anticipated future. Such anticipations must be tuned to the processing delays in the nervous system as well as to the slowness of the body, something that requires constant adaptation of the predictive mechanisms, which in turn require that sensory information be processed at different time-scales.

The target article presents an interesting analysis of visual prediction, and we fully agree with Nijhawan that predictions need to be made at different levels of the sensory-motor processing. However, Nijhawan's article severely underestimates the complexity of anticipation in the sensory-motor system. In particular, it ignores the important question of how an anticipatory system can tune its predictions to internal and external delays as a result of experience. Consider the example of catching a ball. It is not sufficient that "the visual system modifies the perceived position of the ball so that it matches the position of the moving hand," as Nijhawan suggests (sect. 10, para. 3). Instead, such a task involves at least the following components that can be divided into a visual pursuit and a catch component. Even to just visually focus on the ball, its trajectory needs to be anticipated (Balkenius & Johansson 2007). Since visual processing and the eye movements following it are not instantaneous, it is necessary to predict where the ball is right now to be able to fixate it while it is moving. We call this the "anticipated now" because any sensory code that is synchronous with an external state must be based on anticipation.

The predictions resulting in the anticipated now may or may not be correct, but there is no way for the organism to correct these predictions until at a later time, when the true sensory input becomes available. At this time, it is possible to adapt the earlier predictions to the actual sensory input, something that requires that the earlier anticipation, as well as the sensory information used for it, should still be available. This implies that at every moment, the sensory input is used both to anticipate the future and to adapt earlier predictions, but because of processing delays, it cannot be used to code for the current state of the external world. Similarly, eye movements cannot be based on the anticipated now, but must be controlled by the anticipated future. Looking at a moving object therefore requires that the