Use of response time to demonstrate fast use of disparity for depth perception

Benjamin Backus
Graduate Center for Vision Research
SUNY College of Optometry

Is depth perception from binocular disparities—stereopsis—slow or fast? Rapidly changing disparities are perceptually difficult to track, which suggests that stereopsis is generally slow, but the wide-spread belief in the slowness of stereo does not have good empirical support. Classic experiments in which stereo occurred slowly are difficult to interpret. We compared speed-accuracy tradeoff functions (SATFs) between two forced-choice discriminations: one based on stereoscopic depth, and one based on luminance. Unexpectedly, both SATFs deviated from chance levels of accuracy at the same response time—approximately 200 ms—with stereo accuracy increasing, on average, more slowly than luminance accuracy after this initial delay. Thus, the initial processing of disparity for perceived depth took no longer than the initial processing of luminance for perceived brightness (preliminary data were reported in a poster by Baptiste Caziot at AIC 2014). Stereoscopically defined surface slant was similarly apparent soon after stimulus presentation. This finding, that binocular disparities are available early during visual processing, means that depth is perceived quickly, and that disparities may be more widely important for everyday visual function than previously thought.

Comparing a deep convolutional network to the brain: shape tuning and translation invariance.

Authors: Dean Pospisil, Anitha Pasupathy, Wyeth Bair

Convolutional neural nets (CNNs) are currently the best performing general purpose image recognition computer algorithms. Their design is hierarchical, not unlike the neural architecture of the ventral visual pathway in the primate brain, which underlies form perception and object recognition. Of interest is whether these networks, following extensive supervised training, end up performing computations that are similar to those at various stages in the visual hierarchy. Early stage filters in CNNs are often Gabor-like, consistent with V1 physiology, so we asked whether mid-level stages in these networks would show response properties like those found in area V4. Specifically, do CNN units encode a translation-invariant representation of boundary curvature in an object-centered coordinate system, as described by Pasupathy and Connor (2001)? To answer this, we used an implementation of "AlexNet" (Krizhevsky et al., 2012) that was trained to classify the labeled images in the 2012 ImageNet challenge. From units in all layers of AlexNet, we recorded responses to the Pasupathy & Connor shape stimuli (51 simple closed shapes at up to 8 rotations) presented at 100 spatial translations (1 pixel increments). For each unit, we fit the responses to the angular-position and curvature (APC) model simultaneously across all stimulus translations. We found that the later layers of AlexNet had a significant fraction of units well fit by the APC model over translation. The greatest fraction (320/1000) was found in the final, fully-connected layer where each unit represents an object category. It is surprising that a network that was neither built to replicate V4 responses, nor trained on the stimuli used to fit these responses, achieves state of the art performance in replicating V4-like boundary curvature tuning. It is also intriguing that only the final layer had more than 10% of V4-like units, given that this layer is associated with object categorization, whereas V4 is thought to be an earlier stage that represents the local shape of object parts. Our results raise the exciting possibility that similar computations underlie object representation in both biological and artificial visual systems. We believe that artificial networks may provide important insight for guiding neurophysiological investigations, and neurophysiology may offer hints for improving artificial vision systems.

Funding: Google Faculty Research Award. NSF Collaborative Research in Computational Neuroscience Grant IIS-1309725.

Neural mechanism for multiplexed detection of texture and luminance boundaries in the visual cortex

Curtis Baker, Amol Gharat
McGill University

Single neurons in the early visual cortex can respond in a consistent, orientation-selective manner to both luminance- and texture-defined boundaries (form-cue invariance). This was previously thought to arise from a summation inputs from two parallel pathways, a linear filter for luminance boundaries and a filter-rectify-filter (FRF) mechanism for texture boundaries. However an alternative suggestion has been that inputs from subcortical Y-cells, which respond to both kinds of information, might play a key role, particularly in light of our recent finding of non-oriented Y-like cells in visual cortex. We propose a model of how this might happen, based on an asymmetric push-pull combination of On- and Off-centre Y-cell signals.

To test this model in cortical simple-type cells, we employ a general linear model (GLM) in which a neuron's response is a weighted linear sum of simulated On- and Off-centre responses to rapid-fire movies of natural images. The GLM regression is done with a machine learning algorithm (iterative gradient-descent, using early stopping regularization with a hold-back dataset). The estimated weights provide spatial maps of how the cortical neuron sums On- and Off-centre LGN inputs.

For each neuron, the resulting parameters specify space-time receptive field maps of On- and Off-pathway inputs. We find that some neurons show approximately balanced, complementary On- and Off-maps, in agreement with the conventional "push-pull" model by which rectified LGN inputs are combined to provide linear spatial filtering. However other neurons exhibit On- and Off-maps that are quite asymmetrical, with a prevailing dominance of the On- or of the Off-pathway inputs.

Thus a substantial fraction of visual cortex neurons do not act as Gabor-like linear spatial filters, contrary to popular models. For cortical neurons receiving input from Y-cells, this "unbalanced push-pull" arrangement could provide a built-in multiplexing of pseudo-linear responses to luminance boundaries, and also form-cue invariant responses to texture boundaries.

Not Just for Binding Anymore: Using Coupled Neural Oscillators for Maximum Selection, Winner-Take-All Sensory Calculations, and Cortical Hue Opponency Modeling

Vincent A. Billock

Coupled neural oscillators have many uses in neuroscience, but until recently their only application in vision science was in binding theory; sensory neural synchronization seems problematic because synchronization changes both spike frequency and spike timing and may thus distort the underlying sensory information. Alternatively, there may be opportunities to use these distortions to accomplish meaningful sensory transformations. Billock & Tsou (2005 & 2011) suggested that synchronization distortions could mimic a power law transformation and allow the calculation of weighted geometric means. Billlock & Tsou (2014) found that synchronization via excitatory synaptic coupling could model the amplification of sensory information during sensory integration in rattlesnakes and cats. Here I explore two other interesting properties of coupled spiking neurons, modeled using Hugh Wilson's modification of the Hodgkin-Huxley equations. (1) If the excitatory coupling model (Billock & Tsou, 2014) used to model rattlesnake amplification is modified by forcing the neurons to undergo spike rate adaptation, then the neurons still synchronize, but do not amplify. Instead they synchronize at the maximum of the neural spike rates. Such MAX operations have been inferred in various sensory and cortical operations. (2) If spiking neurons are very strongly coupled via inhibitory synapses they can behave like a Winner-Take-All (WTA) system. There are two different WTA-like behaviors seen, depending on synaptic time constants: (i) If the synapse are slow, then true WTA behavior is seen, but it takes a relatively long time for settle into its final (steady state) firing rate. If faster synapses are employed, only one neuron spikes while the other produces an oscillatory membrane potential that is synched to the winning neuron's action potentials. Because there seem to be no good published WTA models for realistic spiking neurons, these new models may be useful. I'm using them to model cortical wavelength opponency using electrophysiological firing rate records from macaques as a starting point. Oscillatory neural coupling may turn out to be as useful for sensory transformation as it is for sensory binding. Supported in part by NSF#1456650.

References

Billock & Tsou (2005) JOSA A, 22, 2289-2298.
Billock & Tsou (2011) Psychological Bulletin, 137, 1-18.
Billock & Tsou (2014) JCNS, 26, 1587-1599.

Title: Binocular Function and Eye Alignment During Infancy and Early Childhood
T. Rowan Candy

Studies of cortical physiology have demonstrated that the development of a fully binocular visual system is dependent on both focused and aligned retinal images. This talk will ask how infants are able to maintain eye alignment while their motor demands are changing as a result of growth of the eye and head, and why some individuals lose alignment. A series of studies of the coordination of oculomotor responses during infancy and early childhood will be discussed and interpreted in the context of the developing visual system.

An infomercial for tracking tasks (in the spirit of George's infomercial on attentional filters).
Larry Cormack

In a basic tracking task, subjects attempt to follow a target moving in a random walk. It turns out that quite a bit can be learned from such tasks. In this talk, I will bore former attendees and confuse new attendees about things we have learned from these sorts of tracking tasks (Mulligan et al., 2013; Bonnen et al., 2015). I will then describe some other things that we are doing with tracking tasks (in various stages of maturity) including 3D tracking, multi-object tracking, and tracking Gabor targets.

Perceptual learning of identification

Barbara Dosher
UC Irvine

Perceptual learning, the improvement of visual task performance with practice or training, has been widely though not exclusively studied in simple two-alternative choice designs. This talk extends the integrated reweighting theory of perceptual learning through augmented Hebbian learning to multi-alternative identification tasks.

Orientation and direction perception in reoriented observers

Frank Durgin

Spatial orientation and angular direction must be specified with respect to some reference frame. Normally, the gravitational vertical, the bodily vertical and the environmental vertical are aligned. Re-orienting observers or the world they observe by 45° or 90° shows that a variety of angular biases tend to stick with the (visually or proprioceptively) perceived world reference frame rather than the bodily reference frame. These observations suggest that if biases in environmental statistics are the source of perceptual biases, at least some of these statistics are encoded with respect to the perceived reference frame rather than with respect to retinotopic cortical orientation coding.

Discriminating Depth Edges

James H. Elder
Centre for Vision Research
York University

People are relatively good at perceiving 3D scene layout even without direct depth cues from stereopsis or motion, and this seems to hinge in part upon an ability to detect depth (occluding) edges and correctly assign figure/ground. This is a non-trivial task, as there are many other types of edge in natural images caused by changes in reflectance and illumination. How does the human visual system discriminate occlusion edges from these other 'distractor' edges?

Prior attempts to answer this question have used human observers to hand-label natural image stimuli in an attempt to uncover the image cues that could drive discrimination. Unfortunately this turns out to be a relatively hard labelling task, and the resulting datasets are thus relatively small and noisy.

Here we consider an alternative approach that takes advantage of the new Southampton-York Natural Scenes (SYNS) Dataset. Employing computer vision methods, we automatically extract a very large dataset of occlusion and non-occlusion edges in natural scenes, and use this to evaluate the role of diverse cues for occlusion edge classification. The results will inform future studies of human edge classification.

Submasking: A Key Factor in Human Pattern Vision

Wilson S. Geisler and Stephen Sebastian

One of the most fundamental natural visual tasks is the detection of specific target objects in the environments that surround us. It long been known that the properties of the background have strong effects on target detectability. The most well-known properties are the luminance, contrast, and similarity of the background to the target. In previous studies, we have shown that these properties have highly lawful effects on detection in natural backgrounds (Sebastian et al., 2016). However, there is another important factor affecting detection in natural backgrounds that has received little or no attention in the masking literature, which has been focused on detection in simpler backgrounds. Namely, in natural backgrounds the properties of the background often vary under the target, and hence the target is not uniformly masked by the background (i.e., some regions of the target are masked/occluded more than others). To begin studying this factor we measured detection thresholds in simple independent-noise backgrounds where the contrast, texture orientation, or luminance can change under the target. In each case, the backgrounds were designed so that a classic matched template detector performs equally well whether or not the background changes under the target. When the background does not change under the target then this classic matched template detector is the optimal detector. However, when the background can change we show that the optimal detector weights each pixel location by its estimated reliability. We find that human performance tracks the performance of the reliability-weighted matched template detector. We demonstrate that humans make use of this same principle when detecting targets in natural backgrounds.

Development of screening echocardiogram for detection of asymptomatic left ventricular dysfunction based on the shortest path in a log-polar representation of images

Irmina Gradus-Pizlo, IU School of Medicine
Kunal Agrawal and Zygmunt Pizlo, Purdue University

We plan to develop a limited automated echocardiogram as a screening tool to identify people who have asymptomatic left ventricular dysfunction (ALVD) and who can potentially benefit from proven therapies to prevent the development of heart failure (HF). HF affects 5.5 million Americans, and the cost of treating symptomatic HF is estimated to be $40 billion a year. In most patients HF develops as a slowly advancing disorder and is a result of progressive heart damage by diseases like diabetes, hypertension and atherosclerosis. Patients are not aware of this problem due to lack of recommended screening and once HF is established, the mortality is 90% in 10 years. There are 60mln patients with risk factors and 30mln with left ventricular dysfunction. Early detection of ALVD could produce significant benefits for patients and for population health.

Echo examinations are considered to be a gold standard in HF diagnosis. In the current format, the echocardiogram has at least 50 views, takes 30 minutes to acquire, 15 minutes to perform measurements and 15 minutes to interpret at an overall cost of $600-1,100. In order to introduce it as a screening tool it needs to be scaled down to 6 views which will provide maximal information, it should take less than 10 minutes to acquire, measurements should be assisted by a computational model, the interpretation should provide diagnosis whether the echo is normal vs abnormal and the cost should be around $100.

The suggestion that 6 views can provide enough information for the binary diagnosis was verified in a study of 100 patients: 50 normal and 50 abnormal. Six expert echo readers achieved near-perfect performance using 6 views, as compared to the diagnosis based on the full echocardiogram. The ground truth about the geometry of the heart was provided by one expert echo-reader in the form of drawings and measurements. The drawings were used as a prior in a computational model that extracts the contours of heart chambers automatically. In the current version, the contour of the left ventricle is estimated by the model as the shortest path in a log-polar representation of the ventricle. This model has already been shown to be perceptually plausible (Kwon et al., 2015) and so, is expected to “see” what the human observer sees.

Alterations of Resting-State Brain States in Parkinson's Disease

Naying He, Jason Langley, Shiyang Chen, and Xiaoping Hu

University of California, Riverside

xhu@engr.ucr.edu

Parkinson's disease (PD) is a neurodegenerative disease whose hallmark is its well-known movement disorder. While previous studies have attempted to examine the motor network with functional activation studies, the results have been mixed, most likely due to the difficulty of performing a well controlled motor task by these patients. Structural MRI, both anatomic and diffusion weighted, have identified subcortical abnormalities associated with PD, resting state connectivity studies have been scarce. In addition, the resting state is now well known to be non-stationary but dynamic, making it harder to characterize in PD patients and controls. In this work, we have applied a recently developed method for ascertaining the dynamics in the resting state to the study of PD. Specifically, we analyzed the resting state data of PD patients and matched controls, respectively, with our method, deriving state patterns and temporal statistics of the states for the two cohorts and compared them between the cohorts. Our results indicate that there are significant differences in both the states and temporal statistics of the states, particularly in the state that involves major areas of the motor network.

Title: Monitoring Stress in a Dynamic Real-World Environment

Speaker: Holly B Jimison, PhD, FACMI, Northeastern University

Abstract:

Stress has been recognized as an important risk factor for many health-related conditions, including cardiovascular disease, diabetes, asthma, sleep disorders, depression and back pain. Stress has also been shown to precipitate unhealthy behaviors, such as smoking or over eating. Dynamic measures of stress can be an important tool for health behavior change interventions. With new wearable sensors for measuring heart rate variability and electrodermal activity, we now have an opportunity to detect stressful events and anticipate the need for just-in-time automated coaching in a home environment. In this talk we will describe the approaches and opportunities for detecting stressful events in real time, as well as recent findings on the effects of alcohol on heart rate variability estimates of stress recovery during sleep.

Path Integration and Illusory Contours: Evidence for an Intermediate Representation in Visual Contour Interpolation

Philip J. Kellman, Susan B. Carrigan and Gennady Erlikhman

Field, Hayes & Hess (1993) showed that certain geometric relations of spatially separated Gabor elements supported detection of the "path" from surrounding noise elements. The relation of path integration to contour interpolation, as in illusory and occluded contours, has remained unclear. The geometric conditions for path detection appear to be the same as those for contour relatability in contour interpolation, but detection of a path of Gabor elements is not accompanied by perception of illusory contours connecting the elements. We hypothesized that path detection accesses an intermediate contour-linking representation that is necessary but not sufficient for perception of interpolated contours. After early contour linking, further scene constraints determine whether interpolated contours will be perceived. We tested these ideas in a series of experiments, using converging objective performance and perceptual report measures. Four display types were used: standard Gabor elements, modified Gabors with centers matching the background, and step-edge versions of both of these display types. We tested path detection using a 2IFC task; participants searched for collinear paths or paths in which elements were misaligned by varying degrees. Illusory contour perception was studied with a magnitude estimation procedure. We hypothesized that all displays would perform similarly in path detection, but illusory contours would be perceived only in the displays in which continuous surface color supported perceived contours in final scene representations. Results showed that: 1) All 4 display types showed the same path detection performance: Path detection declined with misalignment to chance performance by about 15-20 arc min of retinal misalignment; 2) Illusory contours connecting path elements were not perceived with Gabor elements or modified "step-edge" Gabors, but were consistently perceived with Gabors whose centers were modified to match the surround; and 3) For the latter element types, illusory contour judgments showed the same function of decline with element misalignment as did path detection. These results support the idea that an intermediate contour-linking stage is shared by both both path detection and illusory contour formation, but perceiving interpolated contours depends additionally on surface constraints.

Interdisciplinarity of Center-Surround Processing: Neural Mechanism, Perception, and Engineering

With the significant progress in our theoretic understanding on the functionality of biological systems and characteristics of human vision, now it is a great interest for imaging and automated system technologies to benchmark the mechanisms of the visual system that accomplishes accurate and efficient information processing. Here I address the relevant issues under the topic of the most basic visual system principle: center-surround processing. Center-surround processing is recurrently found in the biological system architecture, explains some of the most fundamental perceptual phenomena, and is utilized in image processing algorithms. Computationally, it normalizes the information coded at a certain spatial coordinate (e.g. responses of a cell) given the information at in its vicinity (e.g. average responses of neighboring cell), which reduces the redundancy of information to increase the efficiency of representation. This computational process is termed lateral inhibition and at the lower end of the visual system it shapes temporal and spatial properties of our perception. In the image processing domain, edge detection and color modification algorithms take such computation as the basic principle. All these show that the principle of center-surround processing connects biological, perceptual, and mathematical disciplines and is directly used to solve real-world problems. In other words, a new discovery in one of the disciplines can bring an integrated progress in all the related disciplines, which may have a direct applicational value. I discuss some of such examples as well as future possibilities.

Neural circuits for surface and boundary integration

Tai Sing Lee, Carnegie Mellon University

Figure-ground interpretation is ambiguous based on local information alone. Resolution of local ambiguity requires integration of global contextual information. Such integration can be mediated by a set of horizontal connections in neural circuits that encode the statistical priors on the co-occurrence of edge signals and surface signals in natural scenes. In this talk, I will first provide neurophysiological evidence that shows, in addition to the well-known edge association field among orientation-selective neurons, there exists a disparity association field among disparity-tuned neurons in the primary visual cortex. I will then show the functional connectivity associated with the edge and disparity association fields can be learned using a probabilistic graphical model from natural scene data. Our success in using graphical models to link the natural scene statistics to functional connectivity in V1 neural circuits suggests the relevance of this class of models for conceptualizing computation in the visual cortex.

Bosco Tjan: An ideal scientific role model

Zhong-Lin Lu

Professor Bosco S. Tjan was murdered in his laboratory at USC on December 2, 2016, at the pinnacle of a flourishing academic career. The vision science and cognitive neuroscience community lost a brilliant scientist and incisive commentator. This talk is a tribute to Bosco's life, career, and many important contributions to vision science and cognitive neuroscience.

Title: Human representation of uncertainty

Speaker: Laurence T. Maloney, Psychology & Neural Science, New York University

Abstract: We make decisions continually throughout the day. Some of these decisions are the sort familiar to economists, framed in terms of profit and loss, but others involve selection of actions. Do we cross the street now or wait a bit? In past work my colleagues and I have shown that – to a first approximation -- such motor decisions can be modeled by the same framework of value and probability used in modeling economic decisions. However, as we explored further we found that human representation of probability consistently deviated from the objective probabilities of outcomes. I will review recent work by myself and others concerning how human decision makers represent and use probability and frequency information in carrying out simple motor and visual tasks and then describe more recent experiments that lead us to the unexpected conclusion that – at least in some decision tasks – representation of probability is categorized, not continuous.

Joint work with Shih-Wei Wu (Yang Ming University), Hang Zhang (Peking University), and James Tee (New York University).

Color Algebras

Jeffrey B. Mulligan, NASA Ames Research Center

A *color algebra* refers to a system for computing sums and products of colors, analogous to additive and subtractive color mixtures. We would like it to match the well-defined algebra of spectral functions describing lights and surface reflectances, but an exact correspondence is impossible after the spectra have been projected to a three-dimensional color space, because of *metamerism* – physically different spectra can produce the same color sensation. Metameric spectra are interchangeable for the purposes of addition, but not multiplication, so any color algebra is necessarily an approximation to physical reality. Nevertheless, because the majority of naturally-occurring spectra are well-behaved (e.g., continuous and slowly-varying), color algebras can be formulated that are largely accurate and agree well with human intuition.

Here we explore the family of algebras that result from associating each color with a member of a three-dimensional manifold of spectra. This association can be used to construct a color product, defined as the color of the spectrum of the wavelength-wise product of the spectra associated with the two input colors. The choice of the spectral manifold determines the behavior of the resulting system, and certain special subspaces allow computational efficiencies. The resulting systems can be used to improve computer graphic rendering techniques, and to model various perceptual phenomena such as color constancy.

A New Conceptualization of Human Visual Sensory Memory

Haluk Ogmen[1] and Michael H. Herzog[2]

[1] *University of Denver*

[2] *Ecole Polytechnique Fédérale de Lausanne (EPFL)*

The Atkinson–Shiffrin "modal model" forms the foundation of our understanding of human memory. It consists of three stores: Sensory Memory (SM), whose visual component is called iconic memory, Short-Term Memory (STM; also called working memory, WM), and Long-Term Memory (LTM). Since its inception, shortcomings of all three components of the modal model have been identified. While the theories of STM and LTM underwent significant modifications to address these shortcomings, models of the iconic memory remained largely unchanged: A high capacity but rapidly decaying store whose contents are encoded in retinotopic coordinates, i.e., according to how the stimulus is projected on the retina. The fundamental shortcoming of iconic memory models is that, because contents are encoded in retinotopic coordinates, the iconic memory cannot hold useful information under normal viewing conditions when objects or the subject are in motion. Hence, half-century after its formulation, it remains an unresolved problem whether and how the first stage of the modal model serves any useful function and how subsequent stages of the modal model receive inputs from the environment. We propose a modified model of human visual sensory-memory by introducing an additional component whose reference-frame consists of motion-grouping based coordinates rather than retinotopic coordinates. Data supporting this new model will be reviewed followed by a discussion of how the model offers solutions to the paradoxes of the traditional model of sensory memory.

Encoding of object boundaries in primate area V4

Anitha Pasupathy

Despite decades of vision research we still do not know what neural mechanisms underlie our ability to parse complex visual scenes and perceive the component regions and objects. To address this question, research in my laboratory targets single neurons in primate area V4, an intermediate stage of the ventral visual pathway important for form processing. We study responses of neurons to isolated shapes, partially occluded stimuli and texture patches while animals are engaged in a fixation or shape discrimination task. Results thus far demonstrate
that: i) isolated visual objects are encoded in V4 in terms of the curvature of the bounding contour features; ii) accidental contours formed at the junction between occluded and occluding objects are weakly encoded, consistent with perception, iii) V4 responses to partially occluded shapes may be augmented by feedback from higher cortices, and finally, iv) V4 responses are more strongly driven by object boundaries than textures patches. Taken together, these results support the hypothesis that the V4 representation forms the foundation for the perception of objects in visual scenes.

Combination of Game and Model-Based Approaches to Training Fluid Intelligence

Misha Pavel

Despite significant efforts in training executive functions (EF) that would generalize across a spectrum of tasks, resulting gains (far-transfer, including "fluid intelligence") typically yield only marginal effect sizes. As part of an IARPA funded effort, a team comprising researchers from Honeywell, University of Oxford, Northeastern University, and Harvard Medical School, developed a computer game ("Robot Factory") designed to exercise the participant's EFs, including their combinations, but continuously changing the stimulus mapping and semantics to require the participants to generalize across blocks of trials. The design was motivated by empirical results as well theoretical notions that human information processing can be represented by a Von Neumann-like computer architecture. In this presentation we will describe the tasks comprising the game including the adaptive, individual-tailored progression. We will then discuss several results of our analyses including new algorithms to estimate task difficulty and the dimensionality of the participants' ability. The results to date suggest that the human information processing system is based on distributed rather than centralized processing when performing EF-type of tasks.

Computational cortical model of top-down influences on lightness judgments

Michael E. Rudd, Department of Physiology and Biophysics, University of Washington

Bosco Tjan and I shared common interests in computational models of vision and the relationship between visual perception and cortical circuitry. I will describe a computational model that I have devised to account for quantitative lightness judgments in mostly simple stimuli. The psychophysical data reveal a previously unexpected role of top-down influences on lightness perception, as well as a connection between lightness and object representations in visual cortex. These two research themes overlap with key themes of Bosco's own research: namely, attention and object recognition. In my model, two distinct types of top-down mechanisms--edge classification and spatial windowing--interact to influence lightness judgments. Together, these two mechanisms fully account for individual differences in lightness matches made with simple stimuli. I will conclude my talk with some speculations about how the model computations might be realized in visual cortical circuits and by discussing some open questions about the nature of quantitative transformations at different levels of the visual nervous system that should be of interest to anyone working in the fields of computational visual neuroscience or mathematical psychology.

Imaging the human thalamic reticular nucleus

Keith Schneider, U. Delaware

The thalamic reticular nucleus (TRN) is a thin layer of gamma-aminobutyric acid-releasing cells wrapping the dorsolateral and anterior segments of the thalamus that modulates both thalamocortical and corticothalamic communication in the brain (Jones, 1975). It is divided into modality-specific sectors, each respecting the topographic organization of the associated thalamic and cortical regions (Pinault, 2004), and its position in the brain is ideal for synchronizing the activities of disjoint thalamic and cortical processes (Crabtree and Isaac, 2002; McAlonan et al., 2008).

In this talk I will review our recent human neuroimaging studies of the TRN. We anatomically identify the human TRN using multiple registered and averaged proton density-weighted structural MRI scans and drive its functional activity with a dual phase-encoded stimulus. We characterize the retinotopic and temporal response properties in the visual sector of the TRN and measured an inhibitory relationship with the contralateral lateral geniculate nucleus (LGN). In an independent study, we used the population receptive field (pRF) model to estimate the response properties of individual fMRI voxels, and we were able to resolve the retinotopic maps in the TRN, measuring both the polar angle and eccentricity components, receptive field size and hemodynamic response function delay. In a third pilot study, we have been investigating the connections of the TRN to the auditory (medial geniculate nucleus, MGN) and visual (LGN) thalamic nuclei under conditions of stimulus competition, with the aim of measuring the attentional control properties of the TRN.

The thalamic reticular nucleus is an important structure governing the recurrent interactions between the thalamus and cortex that may provide a substrate for unified perception, and these initial observations provide a basis for further characterizations of the role of the TRN in human perception.

Training effective use of peripheral vision in Macular Degeneration

Aaron Seitz

Macular Degeneration (MD) is a debilitative retinal pathology where following central vision loss, patients must learn to use peripheral vision for everyday tasks that require fine-scale vision, such as reading, writing, and recognizing faces (Kwon, Nandy, Tjan, 2013). A key aspect of plasticity following injury in MD is the development of a preferred retinal locus (PRL) where, after loss of the fovea, an area of the periphery is used for fixations. This PRL takes on functional behavioral previously subserved by the fovea. There exists great variability among MD patients regarding the effective development of PRLs and this represents a unique training opportunity since an effective PRL requires both improved acuity at the PRL and proper integration of the PRL into cognitive control networks. As such classical perceptual learning approaches, typically focusing on low-level learning, require modification to be effective in rehabilitation of MD. In this talk, I describe recent research in the field of perceptual learning detailing how train PRLs in normally functioning individuals and introduce a novel coordinated attention training (CAT) framework designed to promote effective development and use of PRLs in those with MD.

Anne Sereno

Cognition: From Single Cells to the Clinic

How does an organ weighing little more than 1000 grams accomplish feats such as object recognition, accurate reach, attention, and memory? I present research that provides insights into how cognitive processes are represented in the brain and how understanding those processes can help improve intervention in cases of disease or injury. I highlight three different approaches I have taken to tackle these issues: novel neuronal population decoding methods, computer modeling, and tablet-based assessment. (1) Much research suggests a division of visual processing into a ventral stream important for object processing and a dorsal stream key for spatial processing. Yet other studies over the years blur these distinctions. A novel intrinsic population decoding approach reveals important encoding differences at a population level and suggests that both streams combine shape and spatial signals in their own way for different functional goals. (2) Recordings from multiple brain regions reveal an adapted neural response to a repeated stimulus. In a simple model, utilizing a physiologically plausible adaptation mechanism, we are able to predict the valence and time course of reflexive attention as well as generate a passive short-term memory of a specific event. In the model, neuronal adaptation and mutual inhibition play distinct roles in reflexive attention and memory, providing insights into the relation between these two fundamental cognitive processes. (3) Finally, while sports-related injuries in 12-17 year olds are the leading cause of adolescent emergency room visits, the clinical diagnosis of concussion is dependent on subjective evaluation and is vulnerable to bias by both examiner and examinee. I present a novel approach to immediately measure sensorimotor and executive brain function, and I report findings that suggest such a tool may be helpful to improve diagnosis, intervention, management, and outcome in brain injury.

Joan Sereno

Vibrations to Representations: Issues in speech processing

Spoken word recognition requires a transformation from the signal to the brain to ultimately make contact with internal mental representations. This talk will address fundamental issues regarding the role of linguistic representations in speech processing. The speech signal conveys many types of information, including phonological, syntactic, and semantic information about a word. A basic tenet has always been the claim of arbitrariness, referring to the fact that the meaning of a linguistic sign is not in any way predictable from its form, nor is the form dictated by its meaning. Data will be presented that question this traditional distinction between form and meaning as well as claims involving the encapsulation of information across linguistic levels. In addition, the role of variability in the speech signal will be addressed since we know the signal also contains vital indexical information, conveying paralinguistic information about the speaker, including regional background, emotional state, as well as age and gender. To account for speakers' ability to use this information, models of spoken word recognition assume that the lexicon contains numerous exemplar representations, each containing detailed information about the word and the speaker while contrasting models assume that lexical representations are stored in an abstract form in memory after a process of normalization has stripped away the numerous sources of variability in the speech signal. Experimental data will be presented that exploit variability to modify phoneme categories, showing sensitivity to and acquisition of novel, non-native language categories, with concomitant adverse effects on the native language. The present talk will present data and hypotheses about the nature of the mapping and the processing of the ever-present variability.

Margaret Sereno

Neural Encoding of Scene Complexity

Have our visual systems been shaped by the natural environments within which we evolved? Natural forms (e.g., mountains, trees, clouds) are visually complex because they feature fractal patterns that repeat at many levels of magnification. Lower complexity fractals are particularly prevalent in visual scenes. We test a 'visual fluency' model that suggests our visual system might have adapted (with increased processing efficiency) to these lower levels of complexity. Functional magnetic resonance imaging data is presented, demonstrating that fractal complexity determines which levels of the visual system are preferentially activated. Patterns of lower complexity robustly stimulated the dorsal and ventral visual streams, areas involved in processing objects. This prompted us to propose, test, and confirm the hypothesis that pareidolic experience (seeing familiar shapes in noise) could be modulated by fractal complexity, with lower levels inducing greater incidence. These findings indicate that the higher-level regions of our visual systems are exquisitely tuned to the fractal mix of coarse to fine structure present in these abstract lower complexity fractals. In addition, viewing these kinds of fractals can induce object perception, providing an appealing explanation for why we often see familiar shapes in low complexity fractal forms in nature such as cloud and rock formations. These results show that scale invariant structure by itself may differentially drive cortical responses and should be taken into account when studying the functional architecture of the visual system.

Martin Sereno

DNA and language: the origin of symbol-using systems

The boost in cognitive power accompanying the acquisition of language by Pleistocene hominins eventually resulted in humans resurfacing the planet, for better or worse. To get a fresh perspective on the origin of this remarkable faculty, we critically distinguish the origin of a system capable of evolution from the subsequent evolution that system becomes capable of. Human language arose on a substrate of a system already capable of Darwinian evolution; the genetically supported uniquely human ability to learn a language reflects a key contact point between Darwinian evolution and language. Though implemented in brains generated by DNA symbols coding for protein meanings, the second higher-level symbol-using system of language now operates in a world mostly decoupled from Darwinian evolutionary constraints. Examination of the Darwinian evolution of vocal learning in other animals suggests that the fixation of a key prerequisite to language into the human genome may actually have required initially side-stepping the urge to mean itself. But, knowing my background, you might ask, how does the visual system fit in? The pre-existing ability of primates and other animals to assemble a series of glances into a meaningful scene may have provided a second key preadaptation to modern, auditory-symbol-driven language understanding. With different biogeography, a different clade -- e.g., birds (dinosaurs) -- might have gotten there first.

Parallel Paths to Bipedalism, Tool-Making, Dense Brains, and Complex Phonic Signaling and Learning.

Paul Sereno

Why does human cognition, in its fullest development, appear to be a unique evolutionary phenomenon in Metazoan history on Earth and, apparently, on comparable Earth-like planets or more distant celestial objects? The question is examined by considering what we know now about the trajectory to human cognition, starting with what is often regarded as the seminal adaptation toward the human condition versus that in related primates: bipedalism. The order of appearance of tool-making, dense brains, and complex phonic signaling and learning is determined and the effect, or lack thereof, these adaptations had on the survival and diversification of hominids. That adaptive trajectory is compared to its parallel within the other lineage of terrestrial amniotes, the Dinosauria (dinosaurs and living descendants), where bipedalism first appeared 250 million years ago, followed by tool-making, dense brains, and complex phonic signaling and learning—in different order and with a totally different evolutionary outcome. In light of these considerations, human cognition appears to be an extraordinary historically contingent condition, the adaptive value of which over millions of years is open to question.

Sara Sereno

Measures of Visual Words

Modern culture and science is predicated on written communication. With the emergence of writing systems around 3500 BC, the visual comprehension of language is an ability not yet subject to evolutionary pressures. Central to the study of reading is understanding the activation and recognition of individual word meanings and the factors that influence these lexical processes. In this talk, I will present some of my research investigating visual word recognition. In general, my work has explored several variables affecting word recognition, including lower-level factors (e.g., word length, word-initial letter sequences, spelling-sound regularity), word-level factors (e.g., frequency, age of acquisition, concreteness, emotionality, semantic size, gender, ambiguity), and higher-level factors (e.g., syntactic complexity, contextual predictability). I have also devised novel gaze-contingent paradigms such as "fast priming" and "parafoveal magnification" which enhance foveal and parafoveal processing during reading. My approach has been to use measures that are capable of capturing the temporal signatures of word recognition processes, including lexical decision reaction time, eye fixation times during reading, and recording of event-related brain potentials (ERPs) in language tasks. Additive factors can be used in conjunction with these techniques to assess the relative timing of different processes, with interactive or additive patterns suggesting concurrent or separate processing stages, respectively. Establishing the precise temporal course of the perceptual and cognitive contingencies of word recognition is necessary for determining the degree to which processes are driven by bottom-up or top-down mechanisms. Accordingly, such findings inform neurally plausible models of language comprehension.

A systematic comparison between visual cues for boundary detection

Thomas Serre
Cognitive Linguistic & Psychological Sciences Department
Institute for Brain Sciences
Brown University

The detection of object boundaries is a critical first step for many visual processing tasks. Multiple cues (we consider luminance, color, motion and binocular disparity) available in the early visual system may signal object boundaries but little is known about their relative diagnosticity and how to optimally combine them for boundary detection. This study thus aims at understanding how early visual processes inform boundary detection in natural scenes. We collected color binocular video sequences of natural scenes to construct a video database. Each scene was annotated with two full sets of ground-truth contours (one set limited to object boundaries and another set which included all edges). We implemented an integrated computational model of early vision that spans all considered cues, and then assessed their diagnosticity by training machine learning classifiers on individual channels. Color and luminance were found to be most diagnostic while stereo and motion were least. Combining all cues yielded a significant improvement in accuracy beyond that of any cue in isolation. Furthermore, the accuracy of individual cues was found to be a poor predictor of their unique contribution for the combination. This result suggested a complex interaction between cues, which we further quantified using regularization techniques. Our systematic assessment of the accuracy of early vision models for boundary detection together with the resulting annotated video dataset should provide a useful benchmark towards the development of higher-level models of visual processing.

Perceptual Resolution of Color from Ambiguous Neural Representations

Steven K. Shevell and Wei Wang
Department of Psychology and Institute for Mind and Biology
University of Chicago

Our ability to see in the natural world depends on the neural representations of objects. Signals sent from eye to brain are the basis for what we see, but these signals must be transformed from the image-based representation of light on the retina to an object-based representation of edges and surfaces. A challenge for understanding this transformation is the ambiguous nature of the image-based representation from the eye. Textbooks examples demonstrate this ambiguity using a constant retinal image that causes fluctuation between two different bistable percepts (as in the face-or-vase illusion, or a Necker cube that switches between two orientations). Bistable colors also can be experienced with ambiguous chromatic neural representations. Recent experiments (1) generate ambiguous chromatic neural representations that result in perceptual bistablity alternating between two colors, (2) reveal that two or more distinct objects in view, each with its own ambiguous chromatic representation, often have the same color, which reveals that grouping contributes to the resolution of chromatic ambiguity, and (3) show that grouping survives even with unequal temporal properties among the separate ambiguous representations, as predicted by a model of binocularly integrated visual competition.

The Perceived Motion Direction of Equal-Spatial-Frequency Plaid Stimuli

George Sperling, Dantian Liu, and Ling Lin
Department of Cognitive Sciences
University of California, Irvine USA 92697

A plaid motion stimulus consists of two superimposed sinewave gratings moving in independent directions and speeds. Every plaid has a physical interpretation as a single rigid translating frame. Perceptually. plaid component gratings sometimes are perceived to move transparently in different directions, sometimes in the direction of the vector sum of the component velocity vectors, sometimes in the rigid direction, and often in yet other ways. There are two types of plaids. In so-called Type 1 plaids, the rigid and vector sum directions always lie between the two component directions, in Type-2 plaids, the vector sum lies inside but the rigid direction lies outside, making Type-2 plaids useful for discriminating theories.

We varied component-contrasts of Type-2 plaids to produce different contrast ratios and used high temporal frequencies (above 10 Hz) to excluding the third-order mechanism and stimulate exclusively the first-order motion mechanism. Explanations of previous plaid experiments have concentrated on “rigid” versus “velocity vector sum” directions, both of which are independent of the contrasts of their components. With our plaids, 1 cycle/deg gratings within a circular Gaussian window, we find something quite different.

At sufficiently high temporal frequencies, for contrasts of the higher-contrast component ranging from 4 to 32%, the perceived plaid direction was entirely determined by the contrast ratio of the components, independent of their overall contrast. Whenever the component's contrast ratio exceeded 4, perceived direction was simply the direction of the higher-contrast component. For intermediate contrast ratios, all perceived directions between the two components were systematically observed. However, when component temporal frequencies decrease below 10 Hz to enable third-order motion processing, the perceived direction of plaids with equal or near-equal component contrasts deviates increasingly towards and the rigid direction which is not between the components, and is perceived direction of slowest, highest-contrast plaids. The data are encapsulated by a vector-addition model in which the vector length (representing motion strength of a sinewave plaid component) increases approximately as the square of its contrast. The vector strengths derived from the Type-2 plaid experiment predicted within measurement error the perceived direction of plaids in a new experiment in which the angle between the two component gratings of a plaid was arbitrarily varied.

Conclusions: When the first-order motion system is stimulated as the primary mechanism, the foveally perceived motion direction of equal-spatial-frequency plaid stimuli is determined by the relative motion strengths of the components, with strength increasing as a power of grating contrast. When temporal frequency decreases, the slower, third-order motion processing influences perceived plaid motion towards the rigid direction. The velocity vector sum and second-order motion directions are irrelevant.

Title: On the lawful relation between discrimination threshold and perceptual bias

Alan Stocker

Abstract: The perception of a stimulus can be characterized by two fundamental psychophysical measures: how well the stimulus can be discriminated from similar ones (discrimination threshold) and how strongly the perceived stimulus value deviates on average from the true stimulus value (perceptual bias). Historically, it has long been believed that these two measures are independent. I will present compelling evidence that this is not the case and that discrimination threshold and bias are rather directly related. The relation is derived from theoretical considerations of optimal stimulus encoding and decoding, and can be expressed as a surprisingly simple mathematical law. I will demonstrate that the law holds robustly across a wide spectrum of available data from a large variety of psychophysical experiments, and therefore seems to express a fundamental principle of perception.

How to Control Attention and What Attention Controls

John K. Tsotsos, *Thilo Womelsdorf* and *Iuliia Kotseruba*
York University, Toronto, Canada

Although much has been written about visual attention, a broad, mechanistic proposal in the sense of Brown (2014) is yet to appear. That is, most ask "how do humans respond to these stimuli", "what brain regions are active during attentive processes" or "what networks are active," instead of "what mechanisms are necessary to reproduce the essential functions and activity patterns in an attentive system." A mechanistic proposal for attention is our key goal. Its foundation is the Selective Tuning model of visual attention, not the most prominent of theories, but one of a few that is mechanistic and perhaps the only one of sufficient attentional breadth with a variety of experimentally supported predictions.

We begin with the key observation that in a neural network, selection of a region of interest does not solve the network interference problem. Anderson and Van Essen's 1987 "shifter circuits," which allow for dynamic shifts in the relative alignment of input and output arrays without loss of local spatial relationships, hinted at this, but then veered away. The variety of information flow problems inherent in network architectures was described in Tsotsos et al. (1995) and further detailed in Tsotsos (2011). How information flow is impacted by network architecture must be considered for both forward and backward flows. The main forward-flow issues were termed Context, Cross-talk, and Blurring while the primary backward-flow issues were identified as Convergent Recurrence, and Spatial Interpolation. The Selective Tuning model provides for a number of attentional mechanisms that ameliorate each of these but no strategy was specified for how those mechanisms are to be deployed and coordinated. One of the goals of STAR, the *Selective Tuning Reference Architecture*, is to provide an attention executive to fill this function. Although still in early stages of development, important observations are emerging.

The main punch-line of Tsotsos (2011) is that the visual system must involve a general-purpose system with a wide repertoire of behaviours, and whose configuration and particular behaviour is determined dynamically dependent on the visual input, context and goals of the moment. If configuration is dynamic, then timing matters and that temporal structure is determined by several elements. The main determinant is the behaviour specification, and in STAR, this is represented by a *Cognitive Program* (CP - an extension of Ullman's Visual Routines). The visual system has a set of basic functions which form constituent elements of these programs. They are stored in a long-term memory and the current task and environment would provide indexes into the memory to select viable CPs. CPs then are executed - they are essentially algorithms for accomplishing a behaviour - and we propose an attention controller that sets, executes and monitors the temporal structure of specific visual tasks with defined durations, beginnings and ends. Such a temporal structure leads naturally to the concept of *Attentive Cycles* as the substrate for sequential visual algorithms to implement a visual task. We suggest that the temporal structure of attentive cycles becomes evident in discretized and rhythmic psychophysical sampling of visual search arrays, and in rhythmic neuronal activity across brain networks at theta and alpha band activities. Taken together, we suggest that attentive cycles are a common currency for coordinating control signals from multiple levels flexibly even when behaviorally goals change rapidly.

The basic visual functions alluded to require further examination. Clearly, the set would include the variety of eye movements as well as the processing of a specific visual input resulting from its transit from the eyes to recognition and decision centers. We suggest that at least one more mechanism is necessary due to the following hypothesis: the brain's visual hierarchy can encode whatever the eyes see, but it cannot discriminate the elements of what is seen in all situations. Three network characteristics lead to signal corruption and interference across the visual field and thus to possible degradation in conflict resolution power: receptive fields are space-limited, representations of the visual field are spatially coarser in higher levels of processing, and feedforward connections converge in many-to-one mapping everywhere. ST's attentional mechanisms applied in a controlled and coordinated fashion are claimed to ameliorate this problem, attempting to maximize a signal-to-interference-plus-noise ratio via *Attentive Beamforming*.

To return to the title, cognitive programs control attention and attention controls how the elements of the visual system implement those programs, imposing a natural cyclic structure and tuning the visual hierarchy to maximize how receptive it is to what it sees moment-to-moment.

Mitigation of publication bias with behavioral process models

Joachim Vandekerckhove

The reliability of published research findings in psychology has been a topic of rising concern. Publication bias, or treating positive findings differently from negative findings, is a contributing factor to this "crisis of confidence," in that it is likely inflating the number of false alarm effects in the literature. We develop a set of generative behavioral models for the biasing process and compare these using Bayesian methods. A Bayesian model averaging approach using Jeffreys weights mitigates the effects of publication bias and allows for a better estimate of evidence.