On Faces, Bodies and Voices: Multimodal Mechanism of Person Recognition – Conference Abstracts:

Pascal Belin

Aix Marseille University

A Vocal Brain: Cerebral Processing of Voice Information

The human voice carries speech but also a wealth of socially-relevant, speaker-related information. Listeners routinely perceive precious information on the speaker's identity (gender, age), affective state (happy, scared), as well as more subtle cues on perceived personality traits (attractiveness, dominance, etc.), strongly influencing social interactions. Using voice psychoacoustics and neuroimaging techniques, we examine the cerebral processing of person-related information in perceptual and neural voice representations. Results indicate a cerebral architecture of voice cognition sharing many similarities with the cerebral organization of face processing, with the main types of information in voices (identity, affect, speech) processed in interacting, but partly dissociable functional pathways.

Brad Duchaine

Dartmouth University

A cumulative model of developmental prosopagnosia

A number of neural findings have led to suggestions that developmental prosopagnosia (DP) results from a disconnection between posterior face-selective areas and more anterior areas. While it is clear that anterior regions are atypical in DP, abnormalities are also seen in posterior regions in many DPs. I will suggest that a cumulative model is a better characterization of the neural basis of DP than a disconnection model. Because face recognition depends on a hierarchically-organized series of areas moving from posterior to anterior, developmental anomalies affecting posterior areas or their connections are likely to affect representations in anterior areas and possibly the structure of anterior areas as well as the connections between

anterior and posterior areas. In contrast, problems in anterior areas may have limited effects on posterior areas. This scenario suggests an increasing percentage of DPs will have problems as one moves from anterior to posterior and group studies will be more likely to show problems in anterior areas, but in many cases, the problems will have originated in posterior areas.

Roberto Caldara
University of Fribourg, Switzerland
Mapping face processing through human diversity

During the last 40 years, the science of face processing has primarily focused on understanding the processes underlying the identification of faces and the reading of non-verbal facial cues to infer emotional states. Despite the vast amount of research on both topics, the exact nature and the specificity of the processes involved in these critical biological skills are still a matter of ongoing debate. Notably, the literature of face processing has largely considered that these distinctive functional and psychological processes subserve universal determinants and unique putative processes. Recently, however, our work with eye movements has disputed this notion by highlighting fundamental differences in face processing between people from Western and Eastern (China, Korea and Japan) cultures. I will also present data from Western observers that similarly also challenges the universality of face processing, favoring instead the existence of distinct and flexible routes.

Paul Downing
Bangor University
The lateral occipitotemporal cortex in action

Perceiving another person's movements and actions engages some of the same representations, as does the actual performance of the action. In a series of studies, we used fMRI with in order to identify the neural bases of these representations. With multivoxel pattern analyses, we identified regions for which local patterns of activity

are similar for a given action, whether it is seen or performed, yet reliably distinct for different actions. Our findings are only in partial agreement with the standard model that focuses on premotor and parietal regions. In particular, across several studies, we identified visual-motor action-specific representations in the lateral occipitotemporal cortex. This broad region is implicated, across several lines of research, in many functions, including perception of motion, tools, verbs, and human bodies and body parts. I will briefly discuss our attempt to draw these related but distinct threads together.

Hillel Aviezer
The Hebrew University of Jerusalem

Extreme emotional expressions:
Why do perceivers fail to utilize diagnostic information?

Although the distinction between positive and negative facial expressions is assumed to be clear and robust, recent research with intense real-life faces has shown that viewers are unable to reliably differentiate the valence of such expressions (Aviezer, Trope, & Todorov, 2012). Yet, the fact that viewers fail to distinguish these expressions does not in itself testify that the faces are physically identical. In fact, a FACS analysis revealed that smiling and mouth opening were more common among winners than losers, indicating an objective difference in facial activity. In a subsequent experiment we supplied participants with valid or invalid information about objective facial activity and valence but valence ratings remained virtually identical and participants failed to differentiate between positive and negative faces. Finally, we propose that the immunity of participants to objectively useful facial information results from stereotypical (but erroneous) inner representations of extreme positive and negative expression. These results suggest a surprising dissociation between information present in expressions and information used by perceivers.

Winrich Freiwald

Rockefeller University

Taking apart the Neural Machinery of Face Processing

Humans, like all primates, are intensely social beings. But primates are not only highly social, they also possess intricate abilities to understand their social environment. How does the brain mediate these remarkable abilities? Imaging work in humans and monkeys shows that regions within the temporal lobes are specialized to process social information. In my talk I will describe our efforts to understand the neural mechanisms of social perception and related cognitive processes, focusing on faces. After all, person recognition frequently starts with face recognition. The temporal lobes of macaque monkeys contain neural machinery to support face recognition. This machinery is organized in a way that allows us to dissect the neural mechanisms of face processing. It consists of a fixed number of discrete patches of face-selective cortex, which we can readily identify with functional magnetic resonance imaging. These face patches constitute face processing modules, concentrating cells encoding the same complex object category, yet multiple, spatially separate face patches exist, each with a unique functional specialization for faces, which are tightly interconnected to form a face-processing network. I will describe this network and how cells within this network encode faces.

Guido Gainotti

Università Cattolica del Sacro Cuore, Policlinico 'A. Gemelli',Italy

Acquired disorders in familiar people recognition through face and voice

The aim of this presentation will consist in discussing modality-specific and multimodal disorders of familiar people recognition, based on the identification of face and voice. Due to the dominance of the visual modality in multisensory integration, faces have usually been considered as the most important channel for people identification and face recognition disorders ('prosopagnosias') have been considered as the most important (and almost exclusive) disorder of familiar people

recognition. This dominance of faces has led to neglect the reports of patients with (right) anterior temporal lobe (ATL) lesions who showed a multimodal defect in people identification and to consider as instances of 'associative prosopagnosia' almost all patients with right ATL lesions who showed a defect of people identification. A review of the literature has shown, however, that in the majority of these patients the investigation had been restricted to faces (or to a comparison between faces and names), but in the few cases in which the vocal modality had also been considered, voice recognition was as (or more) impaired than face recognition. Since a direct comparison between recognition disorders affecting face and voice is not feasible (because normal subjects are more impaired in voice than in face recognition), it is necessary to construct a test battery in which faces and voices of the same famous people are presented for recognition (familiarity judgement) and identification. Furthermore, discrimination between faces and voices of unknown people will be used to identify perceptual disorders. This test battery should allow to identify patients with 'pure associative prosopagnosia', pure associative phoagnosia' and 'multimodal people identification disorders' and to investigate the neuroanatomical correlates of these disorders.

Galia Avidan
Ben-Gurion University of the Negev
Impairment of the face processing network in congenital prosopagnosia

Converging studies suggest that face processing is mediated not by a single, localized brain area, but rather by the contribution of multiple posterior 'core' and anterior 'extended' regions which constitute a coherent, distributed network. One approach to understanding the properties of this face network is to explore its dysfunction in individuals who have difficulties in face processing in the absence of any obvious brain abnormality and in the presence of intact sensory and intellectual functions (congenital prosopagnosia, CP). Intriguingly, individuals with CP exhibit a seemingly normal pattern of functional magnetic resonance imaging (fMRI) activation profile in the 'core' face system. However, both structural and functional MRI studies have documented impairments in the connectivity patterns between the 'core' and the

'extended' systems, and specifically, connectivity with the anterior temporal cortex. To evaluate the pattern of abnormal connectivity in CP, in the current study, using a novel, fMRI-based brain network procedure, we compared the cortical topology of neural circuits for face processing in participants with CP and matched controls. The anterior temporal cortex served as the major hub for the control participants but not for the CPs. In contrast, the CPs evinced hyper-connectivity in posterior visual regions. While consistent with our previous functional and structural results, these results offer new insights by providing a computational, quantitative framework for assessing network structure and topology in cases of impaired face processing and can be applied to other populations with neurodevelopmental perturbations.

Tzvi Ganel
Ben Gurion Univesity of the Negev
Effects of configural processing on the spatial resolution for face features

Configural processing governs human perception across various domains, including face perception. An established marker of configural face perception is the face inversion effect, in which performance is typically better for upright compared to inverted faces. In a series of experiments, we tested whether configural processing could influence basic visual abilities such as perceptual spatial resolution (i.e., the ability to detect minor visual changes). Face-related perceptual spatial resolution was assessed by measuring the just noticeable difference (JND) to subtle positional changes between specific features in upright and inverted faces. The results revealed robust inversion effect for spatial sensitivity to configural-based changes, such as the distance between the mouth and the nose, or the distance between the eyes and the nose. Critically, spatial resolution for face features within the region of the eyes (e.g., the interocular distance between the eyes) was not affected by inversion, implicating that the eye region operates as a separate ‘gestalt’ unit which is relatively immune to manipulations that would normally hamper configural processing. In an additional set of experiments we manipulated pairs of faces by stretching one of the faces along its vertical or the horizontal plain to create an illusory distortion in the perceived distance between two given features (e.g., between the eyes). This allowed us to measure the

effect of inversion on a perceived distance as well the JNDs to detect minimal difference in the same distance within the same experimental design. While the JND results replicated those found in the previous experiments, with no effects of inversion on JNDs within the region of the eyes, the perceived distance between the eyes was significantly affected by inversion. The findings implicate that configural processing modulates different psychophysical abilities in different manners. Furthermore, they indicate that classic psychophysical methods can be used as robust measures of different properties of configural face processing.

Ida Gobbini
Dartmouth College
"Why and how familiar faces are special"

Recognition of the identity of familiar faces in conditions with poor visibility or over large changes in head angle, lighting and partial occlusion is far more accurate than recognition of unfamiliar faces in similar conditions. We used several psychophysical measures to test if familiar faces are processed in a facilitated way. We also collected data with a visual search paradigm to test if one class of social cues transmitted by faces – perception of the direction of another's attention as conveyed by gaze direction and head orientation – is perceived more rapidly in personally familiar faces than in unfamiliar faces. We found a strong effect of familiarity on the detection of these social cues (mean differences of 101 ms and 69 ms for gaze direction and head orientation, respectively) suggesting that the times to process these signals in familiar faces are approximately twice as fast as the corresponding processing times for unfamiliar faces. In light of the results of the psychophysical experiments we have conducted, we discuss hypotheses on how the organization of the visual system for processing faces is enhanced by personal familiarity.

Isabel Gauthier
Vanderbilt University

Recent advances in the study of holistic processing

Holistic processing has been central to the study of face perception for decades. Faces are processed more holistically then non-face objects, and it can be argued that our understanding of face recognition will not be complete until we understand the cause(s) of this difference. A recent view is that human infants are born with the capacity for holistic face processing (Mckone et al., 2012). An alternative is that holistic processing of faces results from our expertise with faces. This accounts predicts that holistic processing should not be specific to faces, but may be obtained for other categories too. Past studies offering evidence for holistic processing of non-face objects have been criticized on several grounds - I will present a large sample training study with non-face objects that offers further evidence of holistic processing caused by training. Another area that remains debated is the underlying mechanism(s) of holistic processing, including whether they are strictly perceptual or attentional in nature. Our training work illustrates that learning to individuate object parts is sufficient to produce holistic processing. We find that holistic processing is obtained when *both* parts of a test object have a history of attention, even if these parts were not previously seen or learned together. In another study, no holistic processing was obtained for novel faces made of parts with a history of not being diagnostic, suggesting that even holistic processing for faces is much more malleable then would be expected by a strong evolutionary view. These results suggest that attentional control plays an important role in the acquisition of holistic processing.

Kalanit Grill-Spector
Stanford University, USA
The functional neuroanatomy of face perception

How does the functional neuroanatomy of the brain enable rapid and efficient face perception? I will describe recent findings from my lab that elucidate the anatomical and functional features of the face network addressing the following questions: Is there a specialized neural hardware (cytoarchitecture) that is dedicated for face processing? In what way do white matter properties and connections play a role in face perception? What are the functional characteristics of population receptive fields of the face network, and how do their characteristics influence bottom-up and top-down processing of faces? Does the fusiform gyrus play a causal role in face perception? To answer these questions, I will provide data from multimodal measurements including fMRI, diffusion weighted imaging, electrocorticography, and cytoarchitectonics as well as examine how functional and anatomical characteristics vary across typical and atypical populations.

Rafi Malach
Weizmann Institute of Science
"Neuronal correlates of inter-exemplar perceptual distance revealed in intra-cranial recordings of the human visual cortex"

The human perceptual apparatus can estimate similarity and distances of face exemplar with remarkable speed and reliability. However the neuronal mechanism that underlies this perceived "distance" is unknown. In my presentation I will present data, obtained from intra-cranial recordings in patients in the course of clinical diagnostic procedures, while vieweing repeated presentations of different face exemplars. Our results reveal that perceptual distances can be reliably derived by measuring the the neuronal distance between patterns of cortical responses generated by these face exemplars in high order, face-selective sites in the visual cortex. Similar analysis in early visual cortical sites as well as control stimuli- such as inverted faces- failed to show similar distance measures. The results support the

notion that the representaitons of individual face exemplars are implemented by the activity patterns across large groups of face-selective neurons- and similarity or differences between these responses- reflect the perceived similarity and differences between face examplars.

Morris Moscovitch
University of Toronto
Title: Episodic memory, the hippocampus and face recognition.

Few models of face recognition take into account the contribution of episodic memory and, by extension, that of the hippocampus. We present a number of experiments that demonstrate that faces of people associated with episodic memories are recognized better than faces that are merely familiar and these, in turn, are associated with hippocampal activation during face perception. Damage to medial temporal cortex also is associated with loss of categorical perception of faces of familiar people, even when the face is recognized well. The results are interpreted with respect to a model of personal semantics in which both autobiographical and semantic information contribute to recognition and perception.

Alice J. O'Toole
The University of Texas at Dallas
Understanding human bodies and their contribution to person identification

Person recognition in natural viewing conditions often begins at a distance. Because the identity-specific information in bodies is visually coarse, identity perception from the body can be robust and accurate even in suboptimal viewing conditions (e.g., poor illumination, low resolution). Although much is known about face recognition, remarkably little is known about the role of the *body* in person identification. This lack of understanding can be traced to limited data aimed at some fundamental and relevant questions about bodies and body perception. How does body shape vary across individuals? How do we describe bodies? How do we combine information from the face and body for recognition? In this talk, I will begin with a simple

description-based approach to quantifying the information in human body shapes. In this work, people rated the applicability of body shape descriptor terms (e.g., stocky, curvy, built) for a large number of people. Correspondence analysis (Benzécri, 1973) was applied to these data to create a similarity space that contained both the bodies and the descriptor terms. Next, we mapped between the perceptual space and a "physical" similarity space that was created from laser scans of bodies (cf., Anguelov et al. 2005). To do this, we created synthetic model bodies of the original people using "related" dimensions of the perceptual and physical body spaces. We found remarkable parity between the two spaces. Reconstructions of the original bodies in the physical space resembled the bodies on which they were modeled. In the second part of the talk, I will cover experiments comparing state-of-the-art, computer-based *face* recognition systems to human performance. These experiments have pointed to people's use of bodies for identification when the viewing conditions make recognition from the face challenging or impossible. The experiments also indicate that body processing for recognition escapes conscious awareness.

Benzécri, J.-P. (1973). *L'Analyse des Données. Volume II. L'Analyse des Correspondances*. Paris, France: Dunod

Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, & Davis, (2005). SCAPE: Shape Completion and Animation of PEople. *ACM Trans. Graph. (Proc. SIG- GRAPH) 24*, 3, 408–416.

David Pitcher

Laboratory of Brain and Cognition, National Institute of Mental Health, USA

How to (temporarily) break the face perception network: combining TMS and fMRI

Faces are rich sources of social information that simultaneously convey someone's identity, attentional focus, and emotional state. Our visual system is so efficient that, to us, processing this wealth of information seems effortless. Yet the simplest functions, like recognizing your mother or judging her mood, depend on the

interaction of multiple specialized brain regions distributed across cortex. The location of these face regions is well established, but our understanding of the unique functions performed by each region and how these regions interact to facilitate face perception is limited. I address this issue by using transcranial magnetic stimulation (TMS) to transiently disrupt face discrimination in neurologically normal experimental participants. In my talk I will describe a series of experiments that combine TMS with fMRI to measure the effects of the induced transient neural disruption across cortex. Using this virtual lesion approach demonstrates how functionally distinct cortical regions are causally connected in the undamaged brain.

Bruno Rossion
University of Louvain, Belgium

Understanding face perception with fast periodic oddball stimulation: a paradigm shift in cognitive neuroscience

When the human brain is stimulated at a fast periodic frequency rate, it synchronizes its activity exactly to this frequency, leading to periodic responses ("steady-state evoked potentials") recorded by the electroencephalogram (EEG). This fast periodic visual stimulation (FPVS) approach has been recently extended to high-level vision by periodically modulating face identity (Rossion & Boremanse, 2011; Liu-Shuang et al., 2014) or face category (Rossion et al., in press). Here I will summarize recent studies illustrating the unique contributions of this approach in the context of fast oddball periodic stimulation: target images (i.e. oddball stimuli) appear at a fixed rate (i.e., periodically) within a rapid (e.g., 5-6 images/second) train of other stimuli. This stimulation generates a robust differential electrophysiological response at exactly the target category presentation rate (i.e., objectively defined), without relying on subtraction procedures and by eliminating the contribution of low-level cues while preserving naturality of the stimuli. Thanks to this approach, within- and between-face categorization responses can be defined and functionally characterized in a few minutes, in typical adults, clinical and developmental populations, opening new avenues for person recognition research.

Rossion, B. & Boremanse, A. (2011). Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady-state visual-evoked potentials. *Journal of Vision.* 11(2):16, 1–21.

Liu-Shuang, J., Norcia, A.M., Rossion, B. (2014). An objective index of individual face discrimination in the right occipito-temporal cortex by means of fast periodic oddball stimulation. *Neuropsychologia*, 52, 57-72.

Rossion, B., Jacques, C., Torfs, K., Liu-Shuang, J. (in press). Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *Journal of Vision*.

Ruth Kimchi

University of Haifa

Faces as Perceptual Wholes: The Interplay between Featural and Configural Properties in Face Representation and Processing

Adults' expertise in face recognition has been attributed to their ability to engage in holistic processing. Exactly what constitutes holistic processing has remained controversial, however. I will describe three studies that address this issue. The approach taken in these studies is derived from the notion that a face is a multidimensional visual object that has both featural and configural information, and the critical question is how these two types of information interact in object identification, discrimination, or classification. In the first study we systematically manipulated the discriminability of facial features (eyes, nose, and mouth) and examined whether the discriminability of the features predicts the discrimination of faces with similar vs. dissimilar configural information (manipulated by inter-eyes and nose-mouth spacing). The second and third studies examined how featural and configural information interact during face processing in normal observers and in a group of individuals with congenital prosopagnosia (CP), using Garner's speeded classification paradigm. Taken together, the results of these studies provide strong converging evidence that intact upright face processing is holistic, as implied by the

perceptual integrality of configural and featural information. Processing of inverted faces is dominated by featural information, and CPs do not perceive faces holistically, as indicated by the separability of configural and featural information in CP's face processing.

Stefan R. Schweinberger
Friedrich Schiller University, Jena
Speaker Perception from Voice and Face

Although humans use their voice mainly for communicating information about the world, paralinguistic cues in the voice convey rich dynamic information about a speaker´s arousal and emotional state, and extralinguistic cues reflect more stable speaker characteristics including identity, biological sex and social gender, socioeconomic or regional background, and age. Here I discuss how methodological progress in voice analysis, morphing and synthesis techniques has recently promoted research that is directly relevant for current theoretical questions, such as how voices are mentally represented in the human brain. I will report experiments that support the distinction between the recognition of familiar and unfamiliar speakers, and other experiments that illustrate the processes and representational changes that accompany the learning of new voices. I will also describe how specific impairments and individual differences in voice perception could relate to specific brain correlates or to other impairments in high-level auditory perception, such as auditory agnosia. Finally, I consider that voices are produced by speakers who are often visible during communication, and present evidence that shows how speaker perception involves dynamic face-voice integration. Overall, the representation of para- and extralinguistic vocal information plays a major role in person perception and social communication, could be neuronally encoded in a prototype-referenced manner, and is subject to flexible adaptive recalibration as a result of specific perceptual experience.

Alexander Todorov
Princeton University
Data-Driven Methods for Modeling Social Perception of Faces

Research shows that people form instantaneous impressions of other people based on facial appearance; b) agree in these impressions; and c) often act on these impressions. These findings suggest that it should be possible to model social perception of faces. However, reducing a high-level social attribution like trustworthiness to the physical description of the face is far from trivial. A major methodological problem is that the space of possible variables driving social perceptions is infinitely large, thus posing an insurmountable hurdle for conventional approaches. The alternative is data-driven approaches whose objective is to identify quantitative relationships between high-dimensional variables (e.g., visual images) and behaviors (e.g., perceptual decisions) with as little bias as possible. I describe a series of studies using reverse correlation methods based on judgments of randomly generated faces from a statistical, multidimensional face model; a vector space where every face can be represented as a vector in the space. These methods can be used to a) model evaluation of faces on any social dimension (e.g., trustworthiness), and b) to identify the perceptual basis of this evaluation. We can model both face shape and face reflectance and experimentally identify their contributions to social perception. These methods provide an excellent discovery tool for mapping configurations of face features to specific social inferences. The methods also have considerable promise for helping to identify the principles of neural coding of faces.

Shimon Ullman
Weizmann Institute
Title: Atoms of recognition

The human visual system makes highly effective use of limited information: it can recognize not only objects, but severely reduced sub-configurations in terms of size or

resolution. Minimal images are useful for the interpretation of complex scenes but they are also challenging because by their nature they are non-redundant stimuli. I will describe human studies and computer simulations of recognizing minimal images in general, and images coming from human face and body in particular. Comparisons show that humans and existing models are very different in their ability to interpret minimal images. I will discuss implications to the representations used for recognition, brain mechanisms, and algorithms for the interpretation of complex scenes, including human agents.

Patrik Vuilleumier
University Medical Center & Center for Neuroscience and Affective Sciences, University of Geneva

Functional organization of face recognition networks: from visual processing to mentalizing

Although it is well known that face perception involves a highly specific distributed brain network, much still remains unresolved concerning the role of different brain areas within this network in the processing of faces and face related information such as emotion expressions. Moreover, parts of this network overlap with systems engaged by affective and social signals conveyed by non-facial stimuli. This presentation will review recent work from our group using fMRI and DTI to investigate the function and structural interconnections of visual and limbic areas implicated in processing faces, facial expressions, and other facial features. It will also illustrate new approaches based on multivoxel pattern analysis of brain activations, allowing us to decode distinct information contents from areas activated during fMRI. The latter approach suggests that different kinds of facial information are represented in different cortical regions within the temporal and frontal lobe that respond to facial expression, with some of these areas holding higher-level supramodal representations of emotions and mental states.

Lior Wolf
Tel-Aviv University
Web-Scale Training for Face Identification

Scaling machine learning methods to massive datasets has attracted considerable attention in recent years, thanks to easy access to ubiquitous sensing and data from the web. Face recognition is a task of great practical interest for which (i) very large labeled datasets exist, containing billions of images; (ii) the number of classes can reach tens of millions or more; and (iii) complex features are necessary in order to encode subtle differences between subjects, while maintaining invariance to factors such as pose, illumination, and aging. We present an elaborate pipeline that consists of a crucial network compression step followed by a new bootstrapping scheme for selecting a challenging subset of the dataset for efficient training of a higher capacity network. By using this approach, we are able to greatly improve face recognition accuracy on the widely used LFW benchmark. Moreover, as performance on supervised face verification (1:1) benchmarks saturates, we propose to shift the attention of the research community to the unsupervised Probe-Gallery (1:N) identification benchmarks. On this task, we bridge between the literature and the industry, for the first time, by directly comparing with the state of the art Commercially-Off-The-Shelf system and show a sizable leap in performance.

Galit Yovel
Tel-Aviv University
Are faces important for face (person) recognition?

Most research on person recognition has focused on the role of the perceptual information that we extract from static images of faces. In the first half of my talk I will present data, which suggest that person-related conceptual information associated with faces may play more significant role than perceptual information in our superb ability to recognize familiar faces. In the second half of the talk I will highlight the role of dynamic information in person recognition. In particular I will show that dynamic information may be important in the integration of face and body as well as

face and voice information and hence in the generation of a multimodal representation of a person.