Theories of Object Perception

Prince of Wales Hotel, Niagara-on-the-Lake, ON | March 23-24, 2026

Zoom: https://yorku.zoom.us/meeting/register/UQL4Td2EShum6-IzrNPH3A

Monday, Mar 23 (Belfountain Room) ………………………………

8:50 **James Elder & Kevin Lande** | Welcome

9:00 **Gunnar Schmidtmann** | The Role of Curvature in Shape Recognition

9:30 **Phil Kellman** | From Neural Units to Constant Curvature Representations of Contour Shape

10:00 *DISCUSSION*

10:30 *COFFEE*

11:00 **Nick Baker** | Constant Curvature Theories of Shape Representation Align with Human Performance on Shape Recognition Tasks

11:30 **Kevin Lande** | Representing Structure with Structured Representations

12:00 *DISCUSSION*

12:30 *LUNCH (Noble dining room)*

1:30 **Sami Yousif** | Object Perception: A Topological Perspective

2:00 **Vlad Ayzenberg** | Shape is Not Necessary for Naturalistic Object Recognition

2:30 *DISCUSSION*

3:00 *COFFEE*

3:30 **Debate**: Be it resolved that boundaries are more important than the regions they bound.

- *First speaker for the proposition*: Nick Baker
- *First speaker for the opposition*: Vlad Ayzenberg

4:30 *POSTERS & RECEPTION (Three Feathers Room)*

6:30 *DINNER (Noble Dining Room)*

Tuesday, Mar 24 (Belfountain room)

9:00 **Dirk Bernhardt-Walther** | Contours, Not Spatial Frequencies, are the Building Blocks of Object Perception

9:30 **Peter Kohler** | The Role of Symmetry in Object Perception

10:00 *DISCUSSION*

10:30 *COFFEE*

11:00 **Tim Oleskiw** | Toward a Neural Code for Natural Object Shape

11:30 **James Elder** |Whence & Whither Contours?

12:00 *DISCUSSION*

12:30 *LUNCH (Noble dining room)*

1:30 **Anitha Pasupathy** | Mid-level Cortical Representations for Object Segmentation and Perception

2:00 **Kohitij Kar** | What Should a Model of Object Perception Predict?

2:30 *DISCUSSION*

3:00 *COFFEE*

3:30 **Debate**: Be it resolved that deep learning renders analytical models unnecessary.

- *First speaker for the proposition*: Kohitij Kar
- *First speaker for the opposition*: Anitha Pasupathy

4:30 *DISCUSSION:* **Next steps**

5:00 *RECEPTION (Three Feathers Room)*

6:30 *DINNER (Hob Nob Restaurant)*

ABSTRACTS

***Vlad Ayzenberg* | Shape is Not Necessary for Naturalistic Object Recognition**

Shape is considered one of the most important visual cues underlying human perception and cognition. However, despite its importance, the current best computational models of vision, deep neural networks (DNNs), do not use shape information, but instead rely on local visual features. Yet, even without shape representations, DNNs have achieved remarkable performance on many visual tasks, and provide the current best neural description of the human ventral visual pathway—the pathway underlying visual recognition. How can DNNs achieve such performance and neural predictivity without shape representations? One possibility is that, in humans, shape is not as important for everyday object recognition as previously thought, and local features are sufficient. To test this possibility, we presented adults and children (5-years-old) with common everyday objects in (1) a 'shape condition' which removes the local features, but preserves shape by presenting objects as silhouettes, (2) a 'feature condition' which disrupts shape information, but preserves local features by spatially scrambling the image, and (3) a 'natural condition' where the image is shown in their undisrupted state. Across three experiments, we found that performance in the local feature condition was as high, or higher, than in the shape condition, suggesting that local features are sufficient for everyday object recognition. In a second set of experiments, we explored the functional role of shape and tested if it may be important for one-shot category learning. In three more experiments, we found that participants were consistently better at categorizing newly learned objects using their shape than local features. Together these findings shed light on the functional role of shape and local features in familiar object recognition and novel object learning. Moreover, these data may illustrate why DNNs, despite not using shape, are able to nevertheless capture fundamental aspects of human perception.

***Nick Baker* | Constant Curvature Theories of Shape Representation Align with Human Performance on Shape Recognition Tasks**

How does the visual system encode representations of a shape's bounding contour? One theory posits that it segments the contour into regions of similar curvature and represents each region with a single curvature value. I begin by sketching a computational model of how the visual system might segment and encode 2D shapes in terms of partial circles, which we refer to as constant curvature segments. The model makes predictions about which shapes can be encoded as visual representations more easily based on the number of segments composing them. We tested these predictions with a shape recognition task. Shapes were classified by their complexity according to the model of constant curvature shape representation. "Few segment" shapes were composed of 9-15 constant curvature segments while "Many segment" shapes were composed of 24-31 segments according to the segmentation model. We tested observers' sensitivity to shape changes at a range (50-400 ms) of brief exposure durations for both few and many segment shapes. We found a consistent advantage in the same/different task for shapes made up of fewer constant curvature segments. This performance advantage could not be explained by greater physical dissimilarity between few segment shape pairs or by greater dissimilarity according to a battery of shape descriptors from ShapeComp (Morgenstern et al., 2021). How unique are the predictions of constant curvature theory regarding the representational complexity of shapes? We tested this in a follow-up experiment by measuring shape complexity according to maximum a posteriori (MAP) skeletal shape representations (Feldman & Singh, 2006). There was a significant correlation between both measures' predicted complexity. We generated a new set of shapes in which we constrained few and many segment shapes to be matched in terms of MAP skeletal shape complexity to test if constant curvature theory could predict representational complexity over and above what is already predicted by other theories of shape representation. We once more found a reliable advantage for few segment shapes in the same/different task, even when the few and many segment shapes had equally complex skeletal representations. We conclude that constant curvature theory aligns with human performance on shape recognition tasks in ways that cannot be readily explained by other theories of shape representation or similarity.

Dirk Bernhardt-Walther | Contours, Not Spatial Frequencies, are the Building Blocks of Object Perception

Orientation processing is central to visual perception, yet the contributions of different image features in natural scene orientation processing remain unresolved. Here, we present converging behavioral and neural evidence that human orientation judgments in real-world scenes depend more strongly on contours delineating object structure than on orientation energy from filter-based computations. In Study 1, participants judged the average orientation of image patches chosen to maximize the discrepancy between contour-based and filter-based orientation estimates. Judgments consistently matched contour-based estimates, indicating a perceptual bias for object boundaries. In Study 2, we modeled fMRI responses from the Natural Scenes Dataset using three image-computable models: Photo–Steerable Pyramid, Line drawing–Steerable Pyramid, and a Contour-based approach. Models prioritizing contour structure better explained the neural data and identified discrete patterns of orientation preferences. Collectively, these findings show that the human visual system encodes scene orientation by extracting contours, the fundamental components of shapes, over filter-based orientation energy. This challenges prevailing models of visual representation and underscores the importance of boundary information in natural vision.

James Elder | Whence & Whither Contours?

Image contours are central to perception. They signal the boundaries of objects and mark changes in illumination and reflectance. They encode virtually all of the perceptually important information in an image in a convenient, compressed code. Decades of neuroscience and psychophysics research show that the visual system is finely tuned to detect, discriminate and encode contours, and these contours determine our perception of object shape.

However, the problem of extracting complete contours from real images turns out to be quite hard, and we still lack a plausible model. AI models for object recognition, detection and segmentation essentially ignore contours, operating instead on pixels and image patches and relying heavily on semantic information.

In this talk I will argue that the computation of contours is not well-modeled by a convolutional or transformer network. Instead, we propose to model contour inference in two stages. In the first stage, early visual mechanisms are used to identify a probabilistic graph that is fit to the data in the image. In the second stage, message passing is used to extract cycles and paths from this graph corresponding to the contours in the image. Critically, the graph and these messages must reflect not just local relationships but strong global priors.

While this is far from a complete theory, I hope that laying out some of the issues, ideas and components will generate discussion and stimulate further research in this direction.

Kohitij Kar | What Should a Model of Object Perception Predict?

Deep artificial neural networks (ANNs) trained for object recognition have emerged as some of the most powerful computational models of the primate ventral visual stream. These models not only achieve human-level object recognition performance but also predict neural responses in inferior temporal (IT) cortex with surprising accuracy. As a result, ANNs are increasingly used as mechanistic hypotheses about the computations underlying object perception. Yet a critical theoretical question remains: **what level of neural predictivity should we actually expect from a correct model of object perception?** Answering this question requires understanding how similar neural representations are across individual brains. If two monkeys' IT cortices differ substantially, no model can reasonably be expected to predict one brain better than another brain can. Thus, cross-individual comparisons provide a principled reference point for evaluating computational models. To formalize this idea, **we first developed a theoretical framework** that uses ANN representations to simulate how different degrees of shared and idiosyncratic variance across brains should appear in standard model–brain comparison analyses. In this framework, we generate synthetic "brains" by systematically varying the transformations between representational spaces, including identical copies, linear mixtures, nonlinear transformations, and combinations thereof. These simulations reveal how commonly used alignment methods, such as one-to-one neuron matching, and many-to-one regression behave under different biological scenarios and realistic constraints such as limited neuronal sampling. We then **tested these predictions using large-scale neural recordings from the IT cortex of eight macaque monkeys** viewing hundreds of natural images. The results show that IT representations across individuals are neither identical nor fully linearly related. Instead, inter-monkey comparisons reveal a mixture of conserved neurons, linearly transformed components, and nonlinear idiosyncratic responses. These measurements allow us to estimate **inter-monkey ceilings**: empirical upper bounds for evaluating computational models of object perception. Finally, we benchmark leading ANNs against these ceilings. While several models approach the inter-monkey ceiling during the early phase of IT responses, their predictivity drops substantially during later response phases. Interestingly, cross-monkey consistency also declines during these later phases, suggesting that late IT dynamics may reflect increasingly nonlinear or individual-specific computations. Together, these results indicate that object representations in IT are neither perfectly conserved nor entirely idiosyncratic across brains but instead comprise a structured mixture of shared and transformed components. Quantifying this structure provides a principled benchmark for evaluating models of object perception: a successful model should capture the representational components that are conserved across individuals while accounting for the transformations that give rise to individual variability. By grounding model evaluation in cross-brain comparisons, this framework clarifies what current ANN models explain about ventral stream computations, and, equally importantly, what aspects of object perception remain theoretically unresolved.

Phil Kellman | From Neural Units to Constant Curvature Representations of Contour Shape

How does visual processing transform early encodings—transient responses of orientation-sensitive units—into more durable, symbolic representations of object boundaries and shape? Recent research sheds light on how this transition may occur in contour and two-dimensional shape perception. Evidence suggests that an initial symbolic representation of contour shape may consist of segments of constant curvature (Baker, Garrigan, & Kellman, 2021). We hypothesize that these representations arise from banks of curvature filters ("arclets") built from oriented units linked by constant turning angles (Kellman, et al., 2013; c.f., Poirier & Wilson, 2006). Arclets span multiple turning angles and scales; collectively, they extract from smooth contour input a representation consisting of constant curvature segments. Baker & Kellman (2021) developed a computational model supporting the plausibility of the approach and its agreement with psychophysical data. Here, we describe a neurally plausible model that operates directly on images and produces a symbolic encoding of 2D contours and shape in terms of constant curvature segments. Each arclet unit is built from combined outputs of three co-circular, odd-symmetric, Gabor filters. In this initial neural model, we use 6 scales with 6 positive and 6 negative turning angles, plus a zero turning angle, per scale. Filled 2D shapes on homogeneous backgrounds are convolved with each filter type, and activations--normalized relative to maximum possible responses--are found for the best fitting arclet position of each type. Arclets with different turning-angle and scale combinations compete to capture local contour segments into curvature bins spaced to be consistent with human sensitivity to curvature differences. Initial tests indicate that the model outputs agree well with human perception, even for arbitrary shapes lacking any constant-curvature parts. Encoding with curvature filters into bins of scale and turning angle yields a scale-invariant shape code that provides a natural account of the ease with which perceivers detect shape similarity despite differences in size.

***Peter Kohler* | The Role of Symmetry in Object Perception**

TBD

***Kevin Lande* | Representing Structure with Structured Representations**

Objects are richly structured, often consisting of multiple parts, features, and relations. How do perceptual systems represent such structure? One possibility is that complex objects are represented holistically—for example by single "grandmother neurons" (Barlow 1972) or by distributed vectors in high-dimensional representational spaces (Smolensky 1988; Morgenstern et al. 2021; Lockhead 1972). A contrasting possibility is that perceptual processes encode complex objects via *structured representations*: primitive representations of parts and features combine into composite representations of objects according to certain rules or constraints, much as words combine syntactically into sentences (Palmer 1977; Vaziri et al. 2009; Feldman 1999, 2023; Cavanagh 2021; Quilty-Dunn et al. 2023; Hafri et al. 2023; Green 2023; Lande 2021, 2024; Nielsen & Connor 2024; Ravencu et al. in press).

In this talk, I will clarify what structured representations are—and what they are not. I emphasize the many ways representations can in principle be structured, and the many ways such structures may be realized computationally and neurally. I articulate several empirical principles, already implicit in a great deal of vision science, for investigating the structure of perceptual representations. Finally, I will illustrate these principles by summarizing a recent set of studies of shape perception, which suggest that representations of object boundaries are hierarchically structured from basic representations of constant-curvature contour segments (Lande et al. in prep; Baker et al. 2021).

***Tim Oleskiw* | Toward a Neural Code for Natural Object Shape**

Planar shape, or the silhouette contour of a solid body, carries rich information important for object recognition, including both local (curvature) and global shape cues. While neurons selective for shape are found within the intermediate cortical area V4 of primates, it remains unknown how populations of these neurons contribute to our perception of objects in natural environments.

Recently, we have used a unique array of shape stimuli that dissociates local and global cues to identify V4 neurons that preferentially respond to natural shape features. Interestingly, preliminary results from a shape discrimination task indicate that the same natural features cue object recognition in both humans and nonhuman primates.

To investigate a population-level natural object code within the ventral visual pathway, we analyzed simultaneous activity recorded from a 96-channel 'Utah' array implanted in area V4 of a juvenile macaca nemestrina observing natural and synthetic shapes. Linear classifiers trained to identify these stimuli decode natural shapes from our V4 population twice as accurately (Mann-Whitney, $p < .05$) than synthetic shapes that lack natural cues. Furthermore, a correlation analysis suggests that V4 population activity is robust to trial-to-trial variability: decoding synthetic shapes, including those with identical local curvature statistics, is significantly more sensitive to shared neural noise.

We conclude with an analysis of object-recognition networks, demonstrating that when trained on low-pass (blurred) natural images, networks preferentially recruit units tuned to natural shape features to learn a robust object code. Together, our results demonstrate that local and global shape features play a fascinating role in the perceptual integration of natural objects.

Anitha Pasupathy | **Mid-level Cortical Representations for Object Segmentation and Perception**

I am interested in understanding how the primate brain encodes visual stimuli and how these representations underlie our ability to segment and perceive visual objects. In my talk, I will present results from experiments that support the hypothesis of joint encoding of shape and surface characteristics in mid-level stages of visual cortex. Specifically, neurons carry information about both the boundary shape and surface texture of objects, quite unlike artificial networks where units appear to be tuned to either boundary form or surface properties. Such joint encoding in the brain may be critical for segmenting objects from scenes. When it comes to perception, however, new experiments with impoverished stimuli evince an "internal object model" supported by top down signals from higher cognitive stages of processing.

Gunnar Schmidtmann | **The Role of Curvature in Shape Recognition: A Psychophysical Perspective**

Curvature is a pervasive feature of the visual world, present in both natural and man-made objects. Evaluating the curvature of object boundaries and surface contours is therefore a fundamental component of visual processing. From an ecological perspective, analyses of natural scene statistics provided evidence for the prevalence of co-circular contours in our visual environment, which may provide important cues for the visual system.

There is a long tradition of research investigating the role of curvature in shape and object perception. Evidence from psychophysics and neurophysiology suggests that the visual system contains mechanisms that are selectively sensitive to curvature and that these mechanisms may contribute to the formation of object representations.

In this talk, I review behavioural and physiological evidence indicating that curvature plays an important role in shape perception, with particular emphasis on specific curvature points, such as local curvature extrema. I will focus primarily on psychophysical findings while relating them to current physiological knowledge and computational models of shape representation. The aim is to evaluate the extent to which curvature-based representations can account for shape perception and to discuss their limitations within a broader framework of object recognition.

***Sami Yousif* | Object Perception: A Topological Perspective**

An object can be described by its Euclidean properties — its height or width, the angles of its branches, the distances between its parts. An object can also be described in purely relational terms: One could describe certain parts as being on top of, near, or adjacent to other parts. Somewhere in between, topological representations capture coarse relational structure without precise Euclidean detail, offering a relatively efficient, low-dimensional representation of spatial content. I will provide evidence from dozens of unique experiments that topological primitives are spontaneously represented in both perception and memory. In doing so, I'll consider two broader questions. First, to what extent are topological primitives related to the idea of shape skeletons? Second, to what extent are these primitives useful for spatial representation more broadly (beyond the specific case of object perception)?

POSTERS

***Matteo Dunnhofer* | Better, But Not Sufficient: Comparing Video Artificial Neural Networks, and Macaque IT Dynamics**

Feedforward ANNs trained on images remain the dominant models of visual cortex, despite being limited to static inference. In contrast, the primate visual system is a dynamical system operating in a dynamic world. Recent evidence shows that macaque inferior temporal (IT) cortex encodes both object identity and motion velocity. How well do static image–trained artificial neural networks (ANNs) capture IT dynamic computations? Do video-trained ANNs improve upon these static models? We recorded activity from 131 IT sites (two monkeys) while they passively viewed 920 short (300ms) videos of moving objects. We tested 12 feedforward, 2 recurrent, and 13 video-trained ANNs. Feedforward models were unfolded in time by extracting frame-by-frame features to approximate instantaneous, history-independent computations. Video ANNs outperformed other models in decoding motion-direction (400 clips; video ANNs accuracy = 0.69, percent correct; feedforward ANNs accuracy = 0.58; Δ accuracy = 0.109, $p<0.05$) and motion-speed (320 clips; video ANNs accuracy = 0.68; feedforward ANNs accuracy = 0.63; Δ accuracy = 0.049, $p<0.05$) on naturalistic videos, paralleling IT (0.58; 0.57) where late responses carried stronger motion information. In a stress test, we generated 100 appearance-free videos (AFV), where appearance was replaced by random pixels preserving motion trajectories. IT decoders trained on appearance-based responses generalized to above-chance motion decoding for AFV (accuracy = 0.59), whereas all ANNs collapsed to chance-level. We then tested how well ANN features predict neural responses. Feedforward models explained a significant portion of IT variance during early responses (~90–180ms; mean %EV = 62.5%) but declined significantly later (~480–570ms; %EV = 23.1%). Video ANNs showed modest improvements in this late window (Δ%EV = 6.9%; $t(24) = 7.55$; $p<0.01$), suggesting that temporal training aids modeling late-phase IT dynamics. Interestingly, recurrent models performed comparably to feedforward networks, indicating that shallow recurrence alone does not bridge the gap. Together, these findings show that temporal training promotes closer alignment between ANN and IT dynamics, yet video models remain bound to appearance-based temporal cues and fail to capture appearance-invariant motion encoding.

***Sabine Muzellec* | Reverse Predictivity Reveals Common and Unique Representations between Artificial Networks and Primate Visual Cortex**

A major goal in visual neuroscience is to develop computational models that capture the internal representations underlying object perception in the primate brain. Artificial neural networks (ANNs) have been increasingly successful at doing so, achieving strong alignment with neural responses in most areas of the visual cortex. Typically, this alignment is evaluated using forward predictivity (how well model features can predict neural responses). However, this approach does not test whether the representational structure of models is itself recoverable from brain activity.

From a theoretical perspective, if two systems truly implement similar representations for object perception, their mappings should be approximately symmetric: the neural population should predict model features, and model features should predict neural responses. Such symmetry would indicate that both systems occupy a similar representational subspace for encoding visual objects. Conversely, an asymmetric mapping would imply that one system contains representational dimensions that are not present (or not accessible) in the other.

Here we introduce reverse predictivity, a complementary diagnostic that measures how well neural population activity in macaque inferior temporal (IT) cortex predicts the activations of units in ANN models. Using this framework, we uncover a striking asymmetry: many models that appear highly brain-like under standard forward evaluations still contain representational dimensions that cannot be predicted from neural data. In contrast, mappings between neural populations across monkeys are largely symmetric, suggesting that the asymmetry reflects genuine representational mismatch between models and the brain.

Reverse predictivity further allows us to identify ANN units that are "common" with biological vision (units that are predictable from neural activity, generalize across animals, and better relate to behavior) versus units that are unique to models. Together, these results suggest that forward predictivity alone does not guarantee representational equivalence: models may match neural responses while still containing dimensions that are not accessible from neural population activity. Reverse predictivity therefore provides a principled way to characterize common representational subspaces and offers a new diagnostic for guiding the development of models whose internal representations more closely match those supporting object perception in primates.

***Matthias Tangemann* | Linking Motion and Objectness in Humans and Machines**

In recent years, deep neural networks have rapidly approached human visual capabilities through end-to-end training on semantic tasks such as object recognition and vision-language alignment. However, research at the intersection of deep learning and psychophysics has also revealed striking differences to human visual perception, such as a lack of robustness and non-human-like errors. My research aims to improve the alignment of human and machine visual perception with a focus on mid-level vision, and in particular perceptual organization. I will present two recent studies centered around the Gestalt principle of "common fate" and discuss how the close link between motion and objectness can be modeled using DNNs. These studies not only reveal further differences between humans and DNNs related to the perception of moving objects, but also offer a perspective on how insights from vision science can help improve the alignment of human and machine vision.

***Maren Wehrheim* | Facial Expression Discrimination Emerges from Partially Overlapping Neural Subspaces of Detection and Identity**

Facial expressions provide a particularly powerful domain for studying theories of object perception as faces share the same fundamental computational challenges as other visual objects. The visual system must infer stable identity while simultaneously decoding changeable attributes such as pose, illumination, and state: problems that are also important to object recognition more broadly. In the case of faces, identity corresponds to object identity, while facial expressions represent dynamic object attributes analogous to color, material state, or deformation. We show that neural recordings from inferior temporal (IT) cortex support this dual challenge by encoding identity and expression within partially overlapping representational subspaces, indicating that both properties are processed within the same high-level object representation system. Neurons that contribute strongly to facial expression discrimination often, but not exclusively, carry information about facial identity and vice versa. This multiplexed coding architecture demonstrates that changeable and invariant features of faces are integrated within a shared visual representation rather than segregated into separate modules. From this perspective, faces represent an ideal model system for probing how the same visual object-perception system simultaneously supports both stable recognition and flexible interpretation of changing visual states. Our findings suggest that facial expression perception emerges from the same high-level object representation system that supports stable face recognition, making faces a rich model for studying general computational principles of object perception.

QUESTIONS

Meta

1. What are the **main questions** that need to be answered to understand object perception? {KL}
2. What criteria would an **adequate model** of object perception have to satisfy? {KL}
3. To what extent can different models of object perception be **unified**? {KL}
4. What are the broader implications of object perception (and its study) for **society** as a whole? {MW}

Curvature

5. **Sufficiency of curvature**: Is curvature information alone sufficient to account for shape and object perception, or must it be combined with additional cues such as contour topology, part structure, or surface information? {GS}
6. **Encoding principle**: Does the visual system explicitly encode curvature extrema, or does it instead encode generally informative contour locations that frequently coincide with curvature extrema? {GS}
7. **Biological plausibility**: To what extent are current computational models of curvature encoding physiologically plausible, and can they generalize to the complexity of natural objects? {GS}
8. In general, shapes with greater **curvature variance** are predicted to be **representationally more complex** according to constant curvature theory. Are there other theories of shape representation that make the same prediction? {NB}
9. If the visual system encodes constant curvature representations of shape, it likely does so early in the processing stream, before representations of a shape's volumetric structure or articulable parts are encoded. What do you think of it as an **early shape abstraction** upon which other theories of shape representation could be constructed? {NB}

Shape

10. How important is **shape** across natural visual tasks? Given that objects in natural scenes are often partially occluded, how often do we even see the entire shape of an object? {VA}
11. What do we mean when we say "**shape**"? Is there an encompassing description that captures "core" human shape representations? Or is shape comprised of multiple, but potentially orthogonal, components? Are some aspects of shape more fundamental than others? {VA}
12. Why do biological vision systems have a **shape bias**? {DBW}
13. What does the **shape bias** imply for the nature of visual processing along the visual processing hierarchy? {DBW}
14. How well can theories of 2D shape perception extend to, or serve as a basis for theories of, 3D shape perception? {KL}
15. To what extent are current theories of shape representation explaining the perception of *distal object shape* (whether its 3D shape or the shape of its boundary, as oriented in 3D space) vs. the proximal shape of the object's projection on the retinal image? {KL}

Symmetry

16. How does **symmetry** contribute to object perception? {PK}

17. Does the contribution change when **symmetry** is present in the object, but not the retinal image? {PK}

Process

18. Is the **feedforward** response to images in the ventral stream solved (i.e. models reach ceiling)? {KK}
19. Segmentation: How are objects **segmented** from visual scenes? This could precede perception or could result from perceptual organization. {AP}
20. Internal models of object: Thinking about perception as inference, what is the role of **top-down** signals? {AP}

Alignment

21. What is the nature of building up object representations in **biological versus machine** visual systems? {DBW}

22. How are we engaging with the questions of idiosyncratic vs. shared **variance** in behavior and responses across monkeys/humans? {KK}
23. How can a theory of human object perception contribute to the development of artificial (computer) vision systems? {MD/MT}
24. When modeling object perception using DNNs, we should consider emergent object representations in generic vision networks rather than DNNs specifically trained for object recognition? {MT}
25. What does it mean for two systems to share the same representation, and how can we empirically test that claim? {SM}
26. When we say that two systems share representations, are we implicitly assuming predictivity between them? Or can two representations be meaningfully "similar" without being mutually predictable? {SM}

Representation

27. Most research in vision seems to lie on one side or the other of a great divide. On one side, researchers attempt to understand mechanisms sensitive to properties of light, such as neural units in V1 sensitive to oriented contrast in local receptive fields. Activations of these mechanisms are transient and fluctuating (e.g., they change several times a second with eye movements) and have been described as "sub-symbolic." On the other side of the divide, work in middle and high-level vision attempts to understand processes and representations relating to material properties of the world, such as objects, surfaces, and events. How can we bridge this great divide? In other words, what do we know about and how can we understand how perceivers attain **abstract, symbolic representations** of the world from initial encodings by mechanisms sensitive to local properties of light? {PK}
28. Representations obtained through seeing in biological systems are often described as **symbolic and/or abstract**. In AI deep learning systems, there appear to be quite general limitations regarding access to information we might regard as symbolic, abstract, or also relational or configural. What criteria can we point to or find to make these descriptions clear, allowing us to **distinguish what or is not symbolic or abstract, etc.** Or, conversely, are these distinctions not really useful or defensible? {PK}
29. What does the study of representational manifolds/geometry/topology add to our understanding of how we perceive and encode the world? {MW}