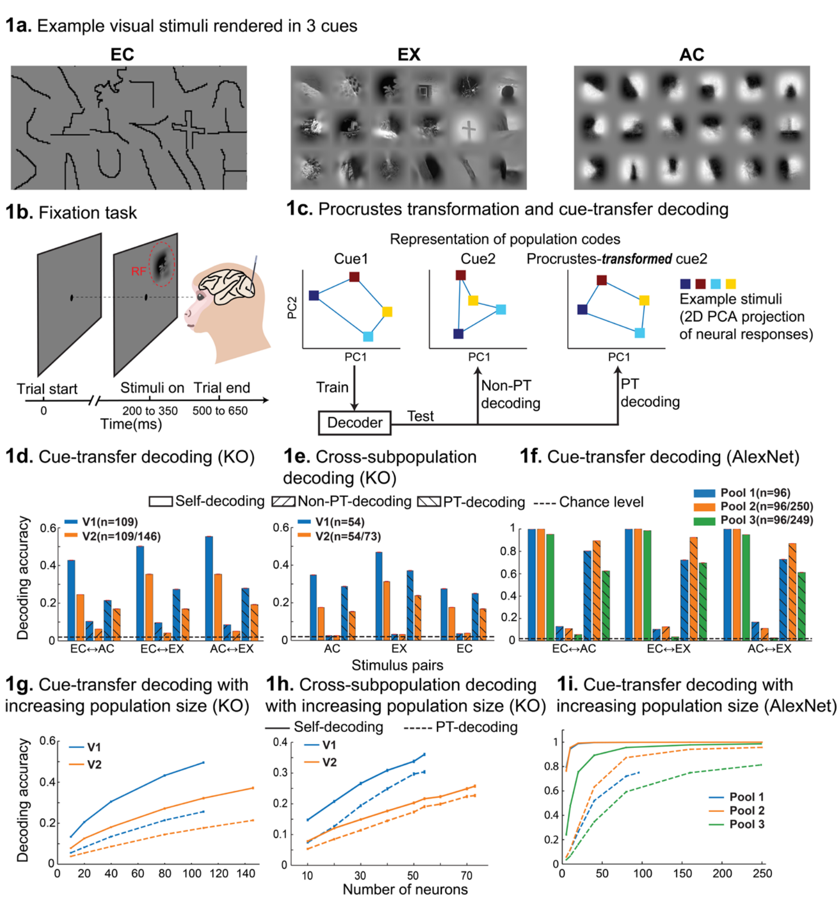
Cue-Invariant Geometric Structure of the Population Codes in Macaque V1 and V2

Summary: Our ability to recognize objects and natural scenes whether depicted in photographs, cartoons, or just line drawings (a.k.a. visual cues) is remarkable. Here, we investigated the representation of a set of surface boundary shapes rendered in different visual cues by macaque V1 and V2 neurons. Specifically, we investigated whether the geometric structure of the population code could support the formation of a cue-invariant representation of abstract surface boundary concepts. A visual concept refers to an abstract feature, such as a boundary or contour, that remains consistent across different types of visual cue rendering. We measured the invariance of the geometric structure across visual cues using a cue-transfer decoding paradigm, i.e., decoding boundary concepts rendered in one cue using a decoder trained on another cue. We found significant cue-transfer decoding when the population codes were aligned via a Procrustes transformation to match each other across cues. The cue-invariance of surface boundary representation was the highest in V1, likely because of its higher resolution in feature dimensions and spatial locations. We observed a similar phenomenon in a model of the ventral visual stream (AlexNet). Cue-invariant boundary representation was degraded in V2 compared to V1, likely due to other invariances, such as translation and rotation invariance, developing along the ventral stream hierarchy. Eliminating the individual neurons' tuning correlations across cues did not adversely affect cue-transfer decoding. The geometric structure was also preserved across different subpopulations of V1 or V2 neurons and between V1 and V2. The stability of the geometric structure increased with the number of neurons participating in the population code. We concluded that despite a trade-off between various forms of invariance along the ventral stream hierarchy, each visual area can optimally build a cue-invariant representation of abstract visual concepts by modifying the geometric structure of its population code.

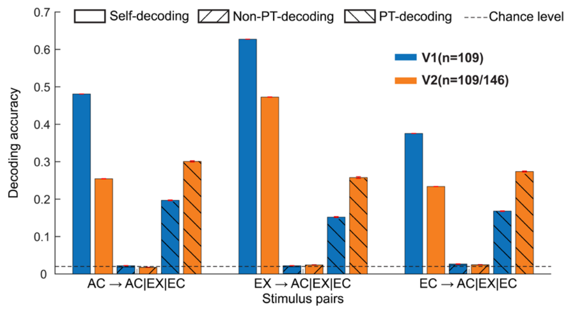
Additional details:

Data/Task: We analyzed V1 and V2 neuronal responses to naturalistic surface boundary stimuli to compare cue-invariance properties across visual areas. Surface boundary elements—such as oriented lines, curves, and junctions—were extracted from natural scenes in the Berkeley segmentation database and grouped into 50 prototype clusters based on contour shapes after segmentation. From each cluster, we generated stimuli rendered in 3 types of cues: (1) Edge Contours (EC), resembling sketch-like object boundary concepts; (2) Example Contours (EX), i.e. the original natural scene patches where the EC concepts come from; and (3) Appearance Contours (AC), capturing the first principal component of natural image patches associated with ECs. **Figure 1a** shows example stimuli rendered in 3 different visual cues, each being a different manifestation of the same underlying abstract surface boundary concepts. **Figure 1b** describes the stimulus presentation sequence while the monkeys performed a fixation task. Stimuli matched average receptive field sizes in V1 and V2. Neuronal activity from 157 V1 and 258 V2 neurons was recorded in two monkeys using multi-contact linear electrodes. Here, we only present data from monkey KO. Both monkeys showed qualitatively similar results.



Methods: To assess the invariance characteristics of population responses, we

2. Inter-area transfer decoding: V1 model decoding V2 and V2 model decoding V1



developed a "cue-transfer" decoding analysis. We trained a linear decoder (support vector machine) on Z-scored neural responses to visual concepts rendered in one cue and tested its ability to decode the same concepts rendered in a different cue (non-PT-decoding, **Figure 1c**). An upper-bound of accuracy was obtained by decoding hold-out trials of the same cue (self-decoding). Cue-transfer decoding was also performed on the Procrustes-transformed population code (PT-decoding, **Figure 1c**). Procrustes analysis found the optimal transformation that aligns the population codes of the two cues (i.e. the optimal rotation; translation and scaling being handled by data Z-scoring). Stimulus-averaged neural responses served as "landmarks" to match the population code of the "source shape" (cue 2) to the "target shape" (cue 1) (**Figure 1c**). Each bar group in **Figures 1d,f** represents the average cue-transfer decoding performance of reciprocal cue pairs (e.g., AC-EC and EC-AC). The horizontal dashed line marks 2% chance-level accuracy for decoding 50 stimuli. We also applied the same analysis to the units' responses of a pre-trained convolutional neural network (AlexNet) to the same stimuli.

Results: **A**. *Cue-invariant geometric structures in V1 and V2 population codes.* **Figure 1d** shows non-PT and PT cue-transfer decoding results for monkey KO's V1 (109 neurons) and V2 (109 neurons sampled from 146 neurons). PT decoding significantly outperformed non-PT decoding in both V1 and V2, achieving roughly half of the upper bound defined by self-decoding accuracy. This indicates that a cue-invariant representation of surface boundary concepts can be achieved through linear transformations (translation, scaling, and rotation) of the population geometry in V1 and V2. Such cue-invariant abstract representation at the population level can, in principle, be achieved downstream through simple neural mechanisms. V1 exhibited better PT decoding than V2, indicating greater cue-invariance in V1's population code. This may be explained by V1 having a more discriminative population code for the boundary visual concepts, which is also reflected by its higher self-decoding accuracy. Both areas performed well above chance, highlighting that, even degraded, V2 exhibited cue-invariance properties, which may reflect a trade-off with other types of invariance, such as translation and rotation invariance, developing along the ventral stream hierarchy. **B**. *Invariant geometric structures across subpopulations within and between cortical areas.* We extended our analysis by performing cue-transfer decoding in two scenarios: (1) Decoding across subpopulations within a cortical area: using a classifier trained on stimuli rendered in one cue using a subset of neurons of the whole population to decode stimuli in the same cue using a different subset of neurons within the same region (**Figure 1e**), and (2) Decoding across areas: using a classifier trained on stimuli rendered in one cue using V1 (resp. V2) neural population to decode stimuli rendered in the same or any other cue using V2 neural population (resp. V1)(**Figure 2**). In all cases, PT decoding significantly improved transfer decoding accuracy compared to non-PT decoding, which remained at chance level due to decoding being applied between different neural populations. This indicates that a linear modification of the population code geometry is a general process that can be applied to build a cue-invariant representation of abstract visual concepts. **C**. *Invariant geometric structures in a model of the ventral visual stream.* We found that the intermediate layers of AlexNet displayed high PT decoding accuracy (**Figure 1f**), with pool 2 being stronger than pool 3. These results are remarkably similar to the neural data**,** indicating that the geometric structure of the population codes in a model of the ventral visual stream could also be exploited to build a cue-invariant representation of abstract visual concepts. **D**. *Higher cue-transfer decoding accuracy with increased population size.* For both neuronal and AlexNet data, cue-transfer decoding accuracy increased as the number of neurons grew, regardless of the decoding scenario, and started to saturate. (**Figures 1g-i**, dashed lines represent the average PT decoding accuracy across all cue pairs). This result shows that the geometric structure of the population code becomes more stable as the population size increases. Overall, we concluded that each visual region, V1 and V2, as well as models of the ventral visual stream, can take advantage of the geometric structure of their population codes, using simple downstream mechanisms, to build a cue-invariant representation of abstract visual concepts.