

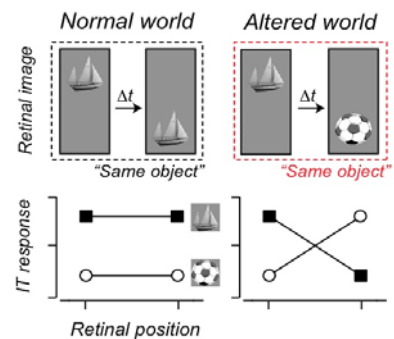
# Natural experience drives online learning of tolerant object representations in visual cortex

Nuo Li and James J DiCarlo

McGovern Institute for Brain Research, MIT

Object recognition is computationally challenging because each object produces a myriad of retinal images. Yet the visual system somehow solves it effortlessly. Neuronal responses at the top of the primate ventral visual stream (inferior temporal cortex; IT) have a key response property that likely underlies this ability -- they are selective among visual objects, yet tolerant to changes in object position, size, pose, lighting, etc. How this tolerant selectivity is constructed remains a fundamental mystery. One possibility is that the visual system builds that tolerance via the spatiotemporal statistics of natural visual experience. Because objects are typically present for relatively long time intervals, while object motion or viewer motion (e.g. eye movements) cause rapid changes in each object's retinal image, the ventral visual stream could construct tolerance by associating neuronal representations that occur closely in time.

If this hypothesis is correct, then we might create "incorrect" tolerance by targeted manipulation of these spatiotemporal statistics. Specifically, if we engineered an altered visual world in which some objects consistently changed identity across retinal position then, following sufficient exposure to this world, the visual system might incorrectly associate the representations of those objects at those positions. The main prediction is that individual IT neurons would lose their normal position-tolerance (i.e. object preference maintained across retinal position), and would instead tend to prefer one object at one position, and another object at the other position (see figure).



We monitored single IT neurons' position-tolerance in two monkeys while they visually explored our altered visual world. We used real-time eye tracking to present visual objects at controlled retinal positions during free viewing: as the animal saccaded toward a specific object (A), it was consistently replaced by another object (B). This manipulation caused the image of object A at a peripheral retinal position ("swapped") to be consistently temporally associated with the image of object B on the fovea. Remarkably, while each animal explored this altered world, its IT neurons gradually began to reverse their object preferences at the swapped position, exactly as predicted. This effect continued to get larger for as long as we could hold neurons (~1 hour), it was specific for object position (counterbalanced across neurons) and object identity, and it cannot be explained by adaptation.

We have previously found that similar manipulations of experience produce changes in the position-tolerance of human object perception [1]. Taken together, our results suggest that the ventral visual stream acquires and maintains a tolerant object representation via the spatiotemporal statistics of natural visual experience, without external supervision. The relatively fast time-scale of this unsupervised learning opens the door to rapid advances in characterizing the crucial spatiotemporal image statistics, understanding other types of tolerance (e.g. size, pose), and ultimately connecting a central cognitive ability -- tolerant object recognition -- to cellular and molecular plasticity mechanisms.

## References

[1] 'Breaking' position-invariant object recognition. Cox DD, Meier P, Oertelt N, and DiCarlo JJ, *Nature Neuroscience* 8:1145-1147, 2005.