

HUMAN HIERARCHY OF TASKS AND ACTIVE SPATIAL PERCEPTION



Andrew Glennerster

Outline

- Evidence against 3D reconstruction
 - some briefly and
 - two examples in more detail
- What does the brain do instead?
 - a 2½-D sketch as ‘base camp’ for different tasks
 - could be implemented as a policy network
- Tomorrow
 - more on hierarchies of tasks
 - a different set of basis vectors for feature learning



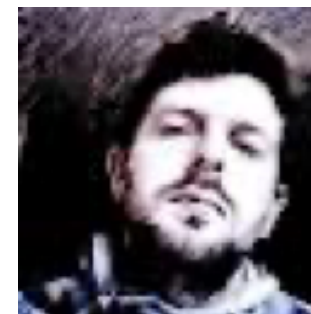
Jenny Vuong



Alex Murry



Luise Gootjes-Dreesbach



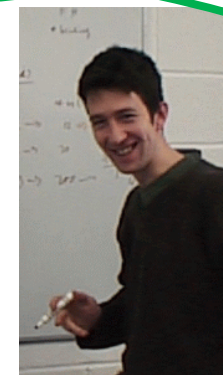
Peter Scarfe



James Stazicker



Miles Hansard



Andrew Fitzgibbon

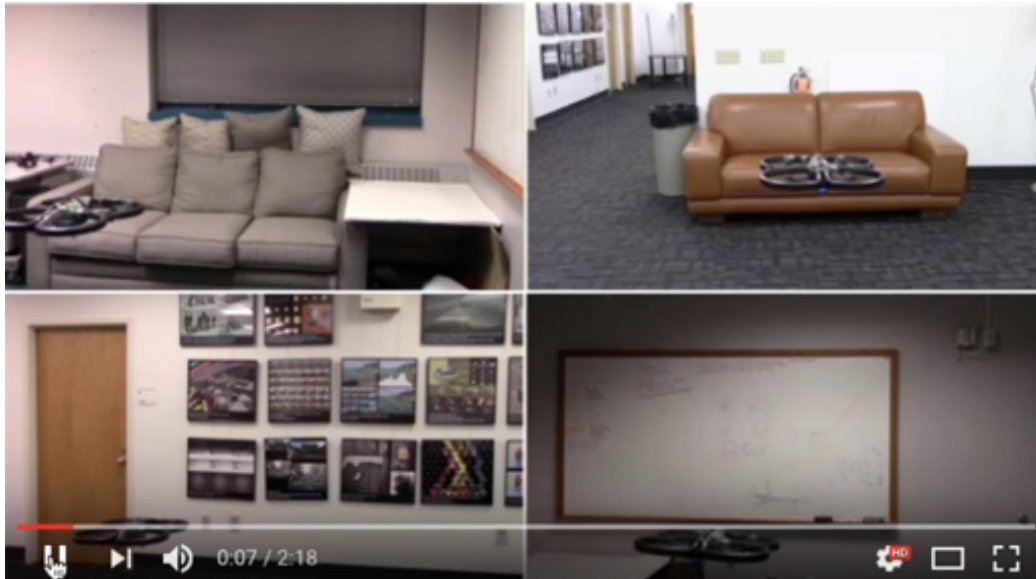
Microsoft
Research

EPSRC

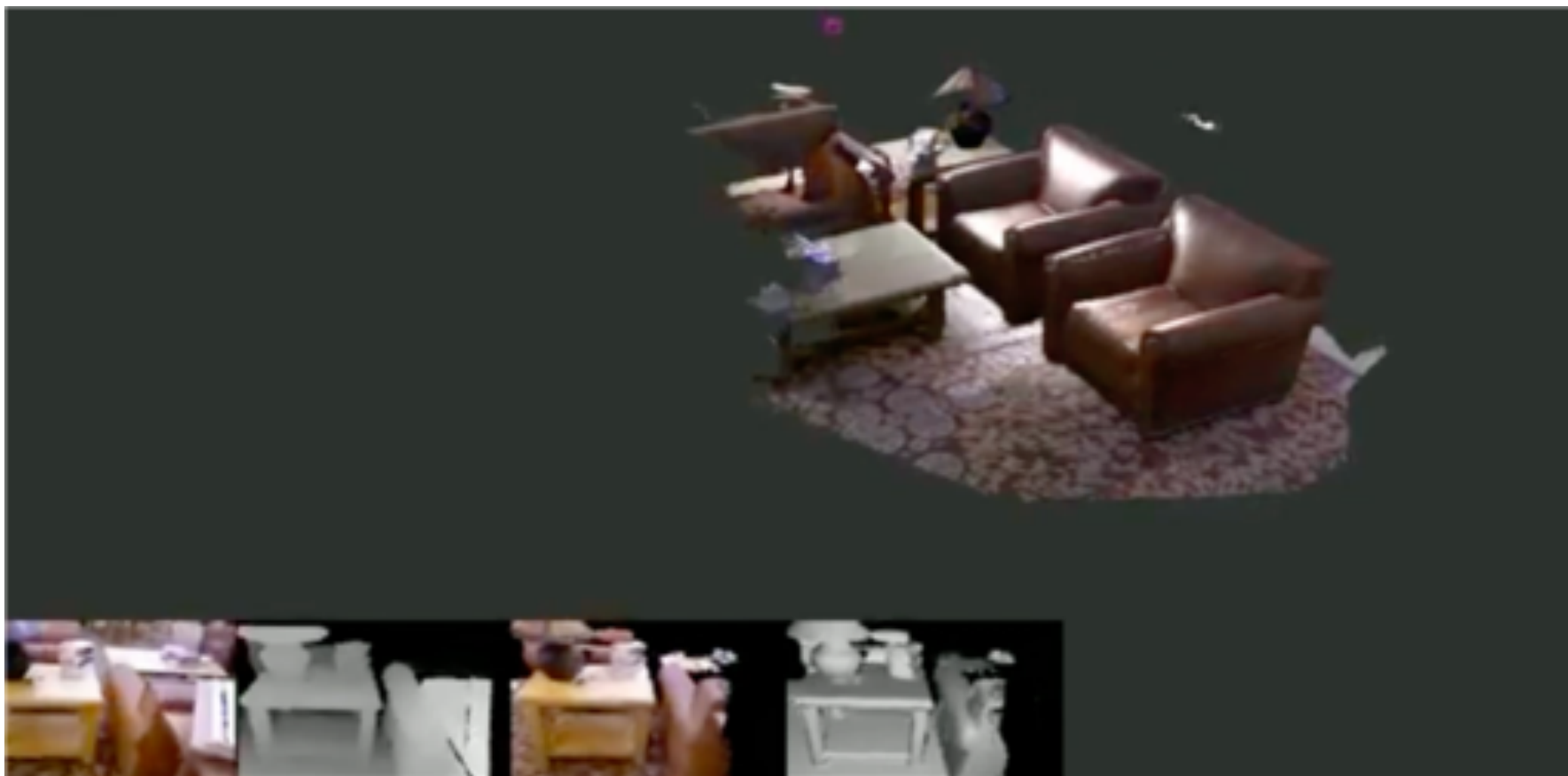
[dstl]

Learning

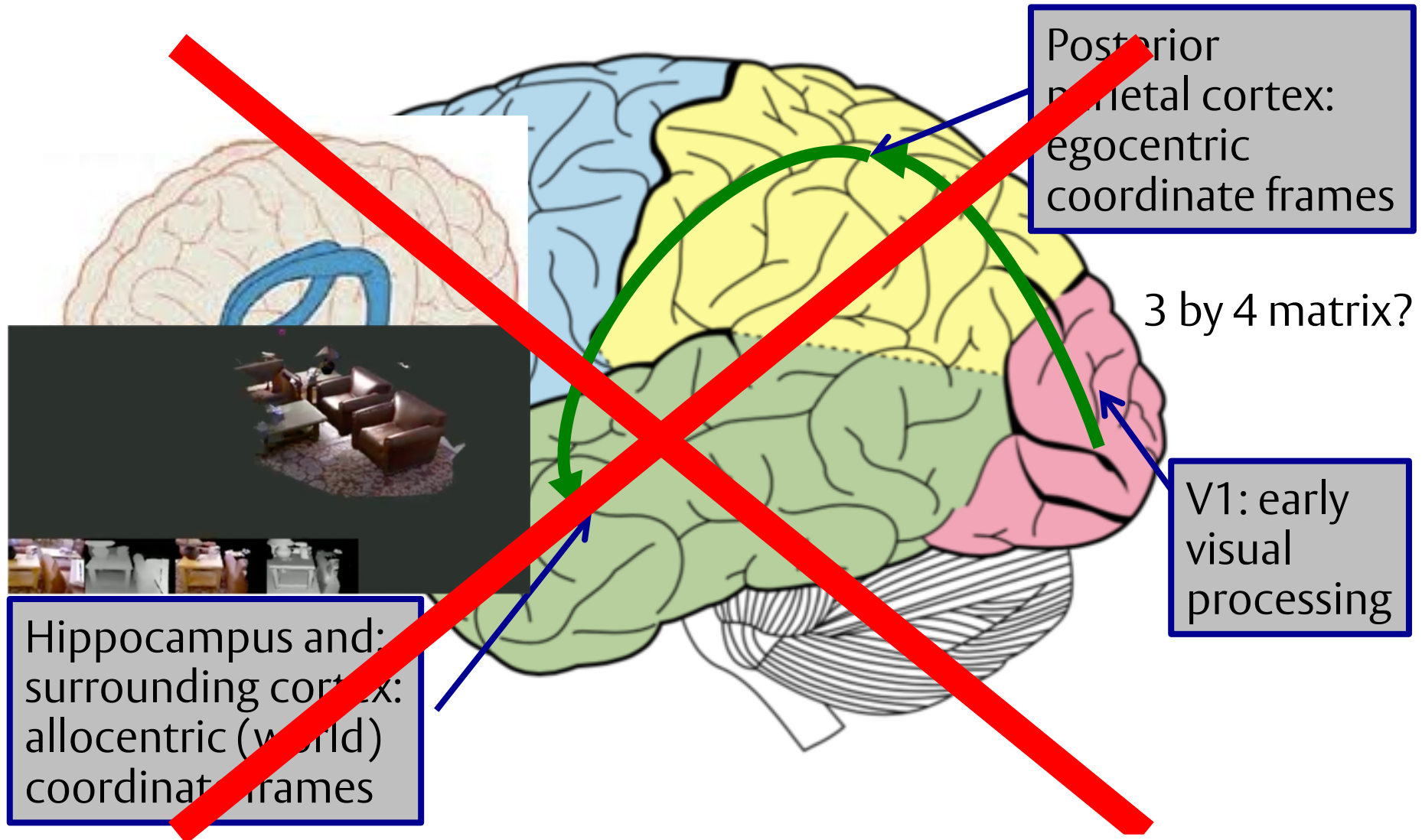
Trained



- Is this all we need (a set of learned policies)?



Current hypothesis



Psychophysical evidence against 3D reconstruction

Hierarchical

Size constancy:

$$h_1 = h_2$$

Depth constancy:

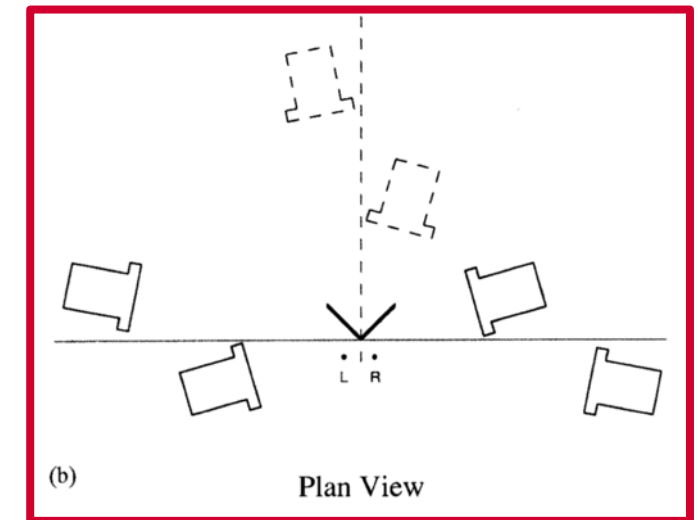
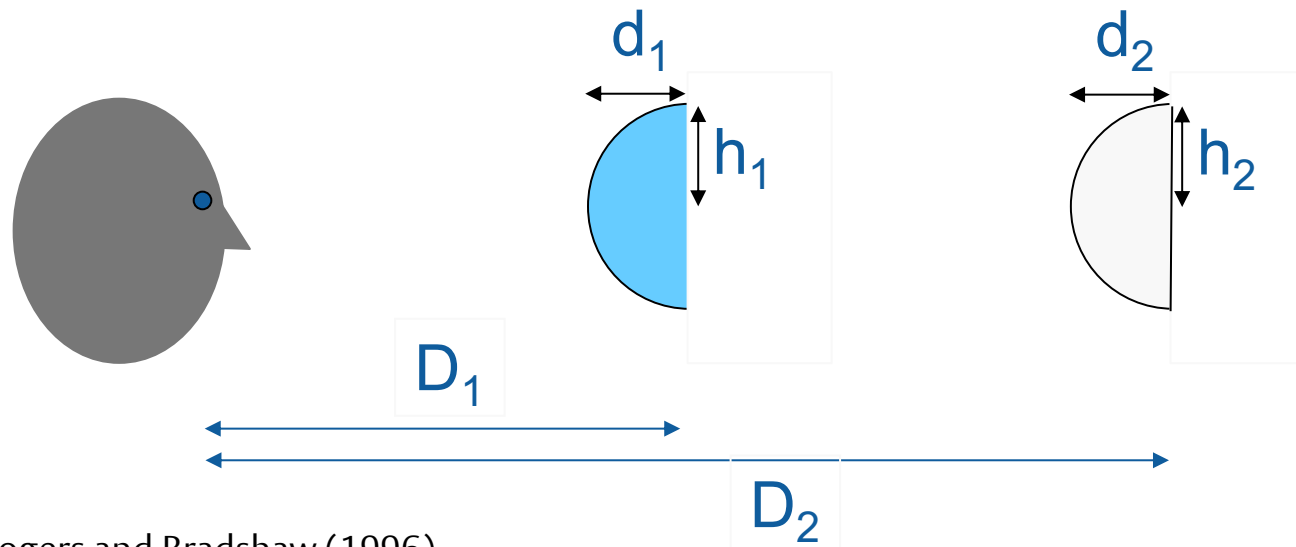
$$d_1 = d_2$$

Depth-to-height ratio:

$$d_1/h_1 \neq d_2/h_2$$

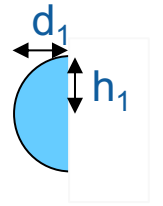


Inconsistent



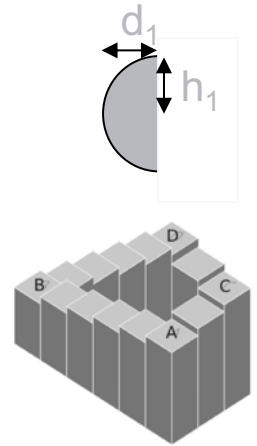
Psychophysical evidence against 3D reconstruction

- Shape judgements depend on the task
 - Glennerster *et al* (1996)

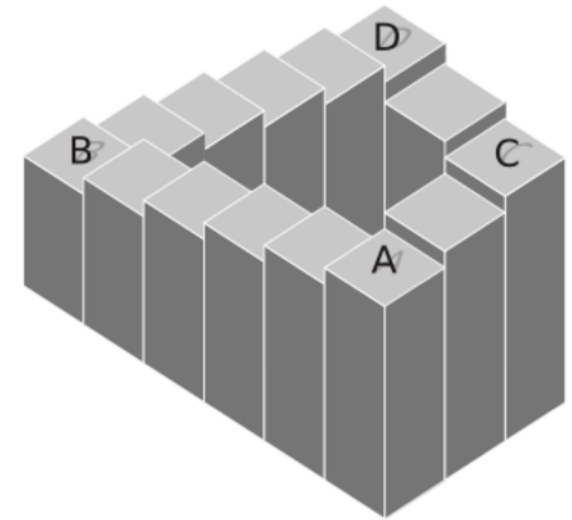
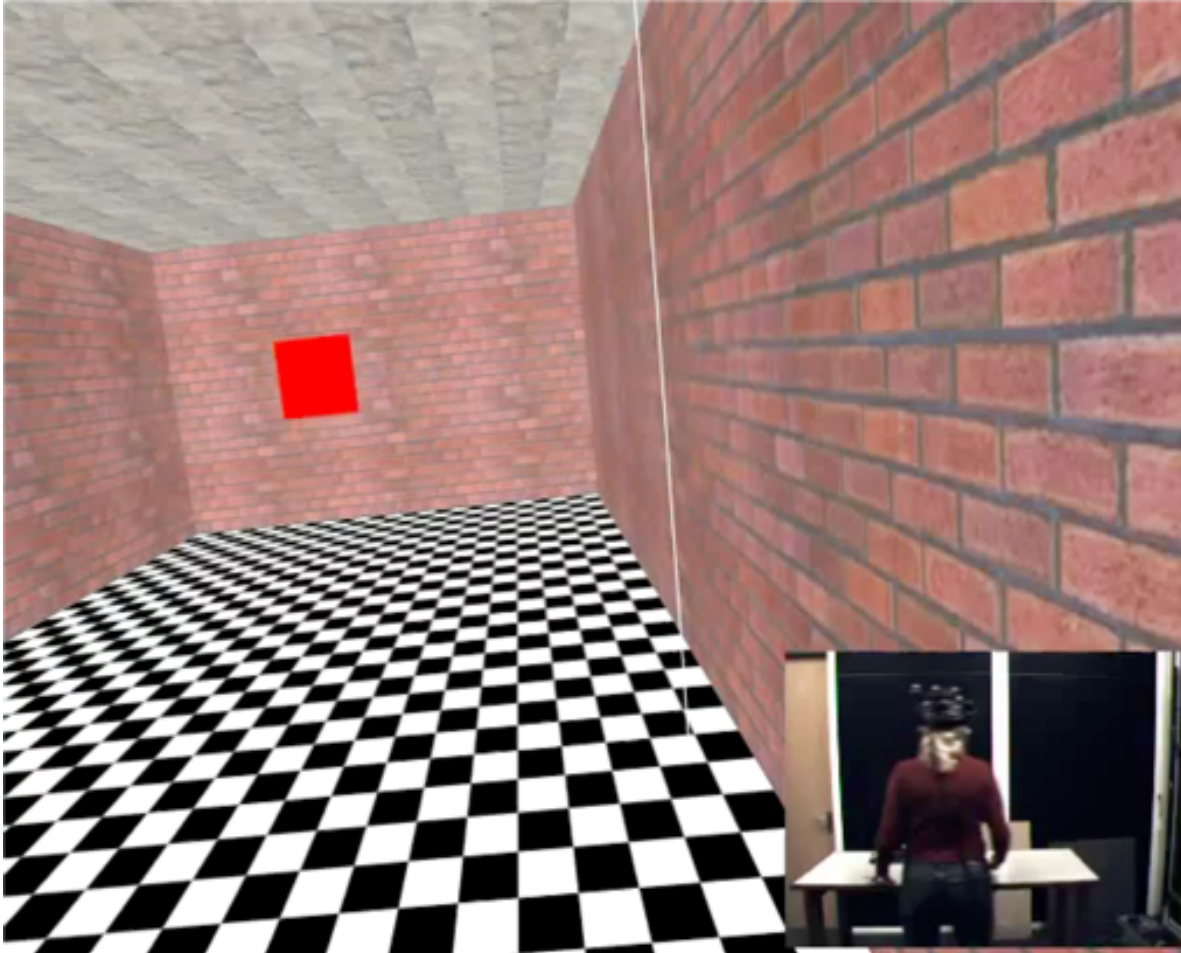


Psychophysical evidence against 3D reconstruction

- Shape judgements depend on the task
 - Glennerster *et al* (1996)
- Intransitivity of depth relations ($A > B > D$ but $A < C < D$)
 - Svarverud *et al* (2012)

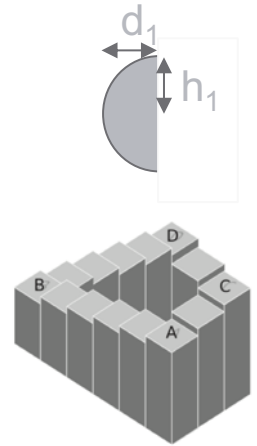


Psychophysical evidence against 3D reconstruction



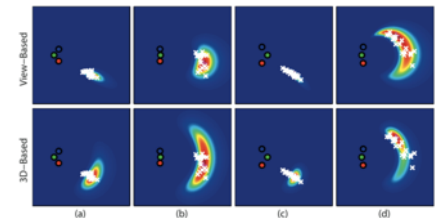
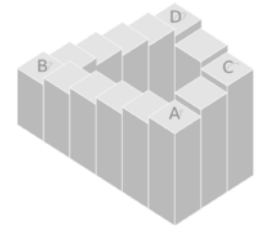
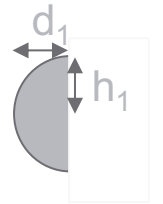
Psychophysical evidence against 3D reconstruction

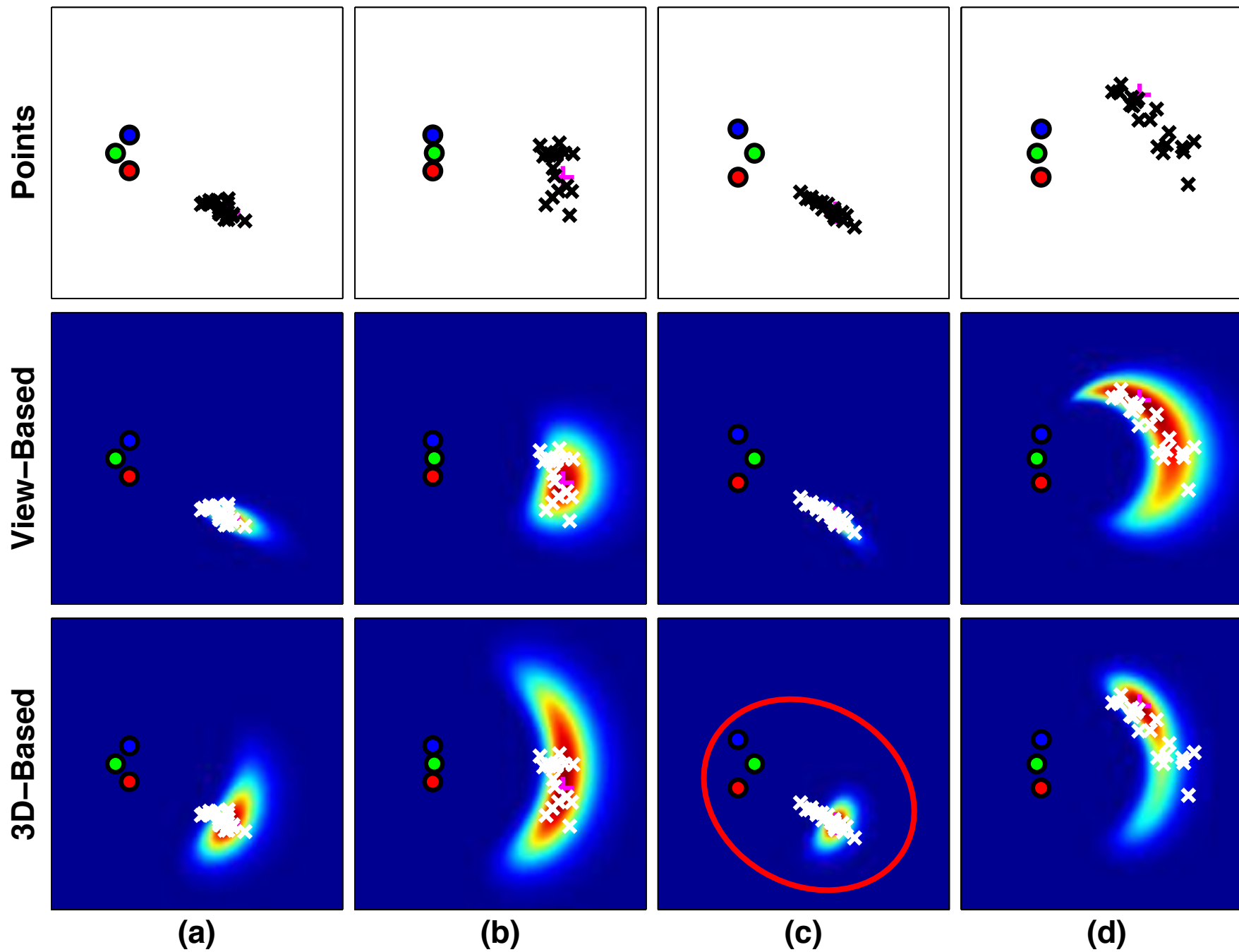
- Shape judgements depend on the task
 - Glennerster *et al* (1996)
- Intransitivity of depth relations ($A > B > D$ but $A < C < D$)
 - Svarverud *et al* (2012)



Psychophysical evidence against 3D reconstruction

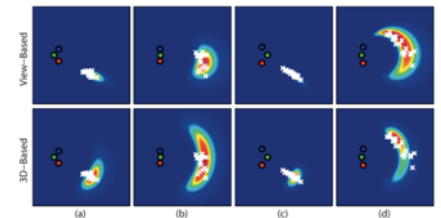
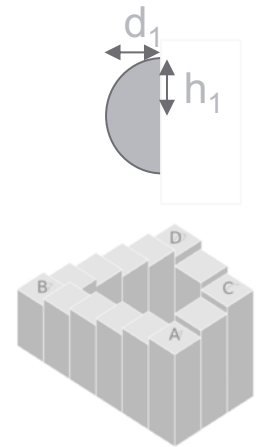
- Shape judgements depend on the task
 - Glennerster *et al* (1996)
- Intransitivity of depth relations ($A > B > D$ but $A < C < D$)
 - Svarverud *et al* (2012)
- Homing errors are better described by a view-based model than 3D reconstruction
 - Gootjes-Dreesbach, Lyndsey Pickup, *et al* (2017)





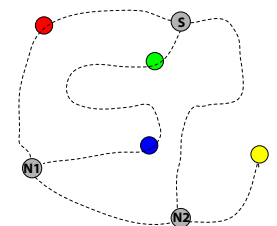
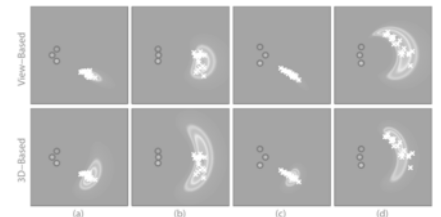
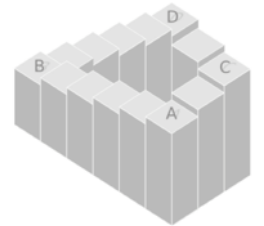
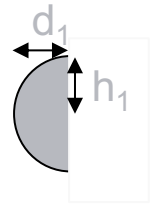
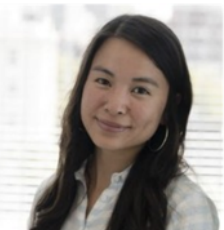
Psychophysical evidence against 3D reconstruction

- Shape judgements depend on the task
 - Glennerster *et al* (1996)
- Intransitivity of depth relations ($A > B > D$ but $A < C < D$)
 - Svarverud *et al* (2012)
- Homing errors are better described by a view-based model than 3D reconstruction
 - Gootjes-Dreesbach, Lyndsey Pickup, *et al* (2017)



Psychophysical evidence against 3D reconstruction

- Shape judgements depend on the task
 - Glennerster *et al* (1996)
- Intransitivity of depth relations ($A > B > D$ but $A < C < D$)
 - Svarverud *et al* (2012)
- Homing errors are better described by a view-based model than 3D reconstruction
 - Gootjes-Dreesbach, Lyndsey Pickup, *et al* (2017)
- Spatial updating is biased in a way that is inconsistent with 3D reconstruction
 - Vuong *et al* (submitted); Murry and Glennerster (2018) ... in more detail



Can we update the visual direction of unseen objects as we move?



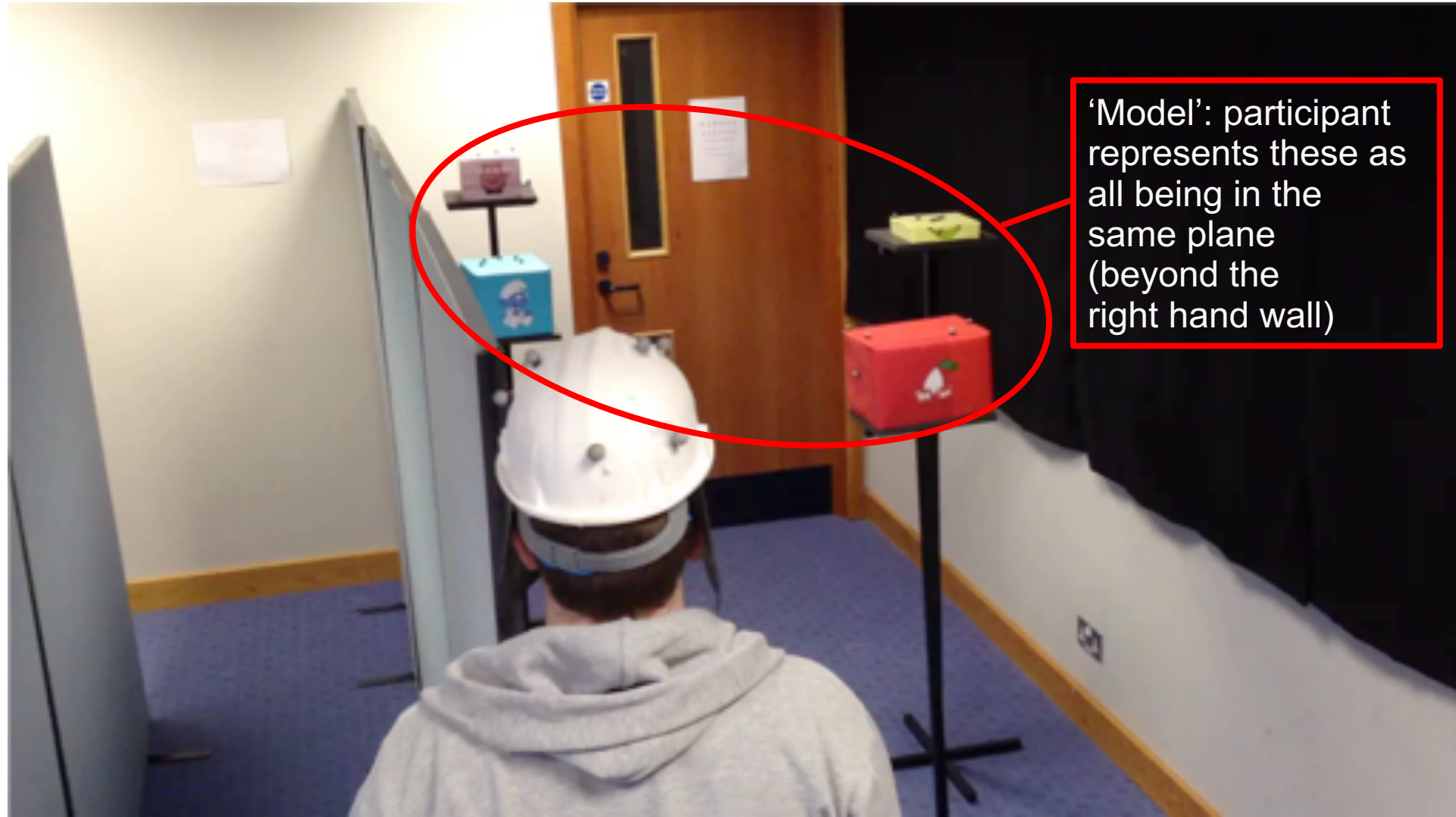
Jenny Vuong



Can we update the visual direction of unseen objects as we move?



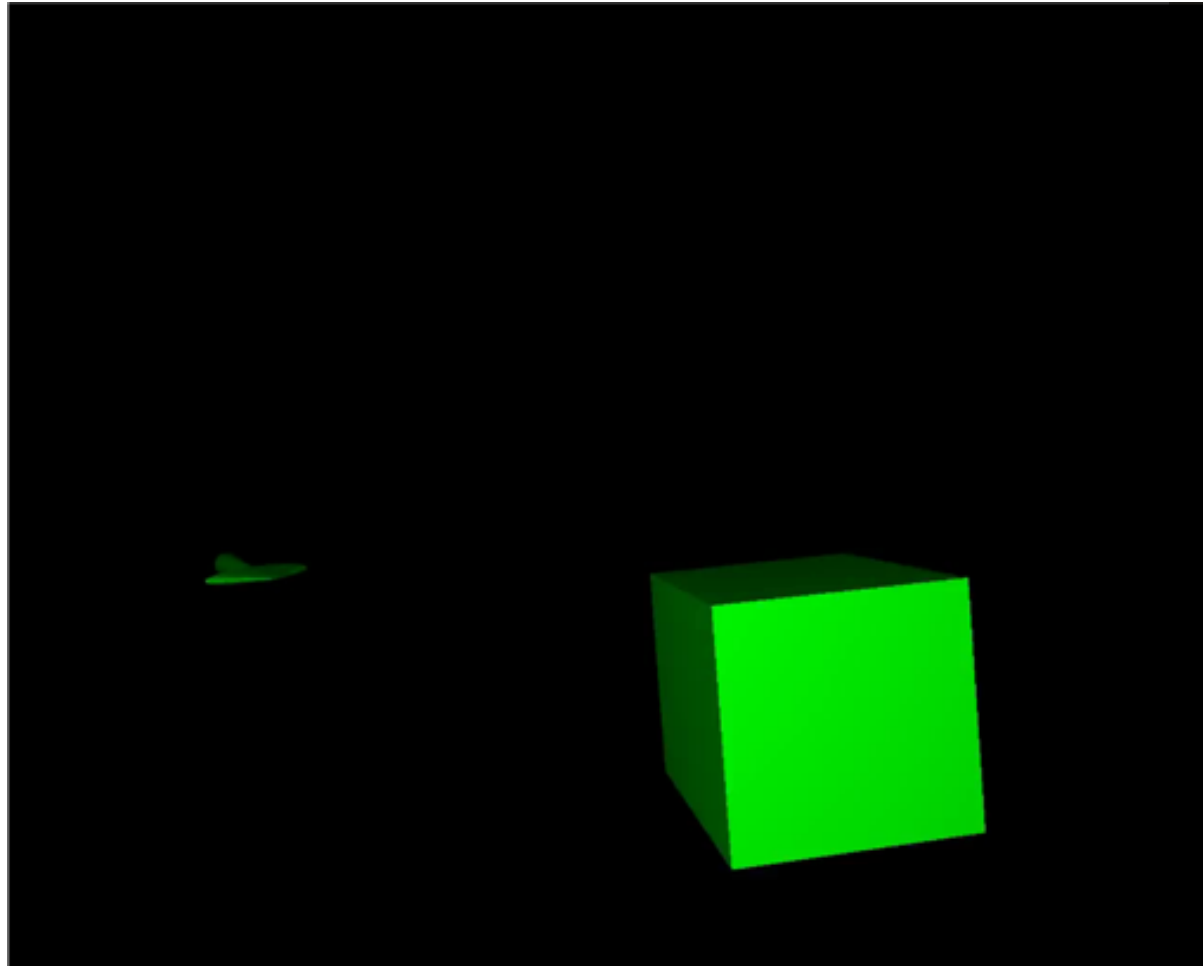
Jenny Vuong



Can we update the visual direction of unseen objects as we move?

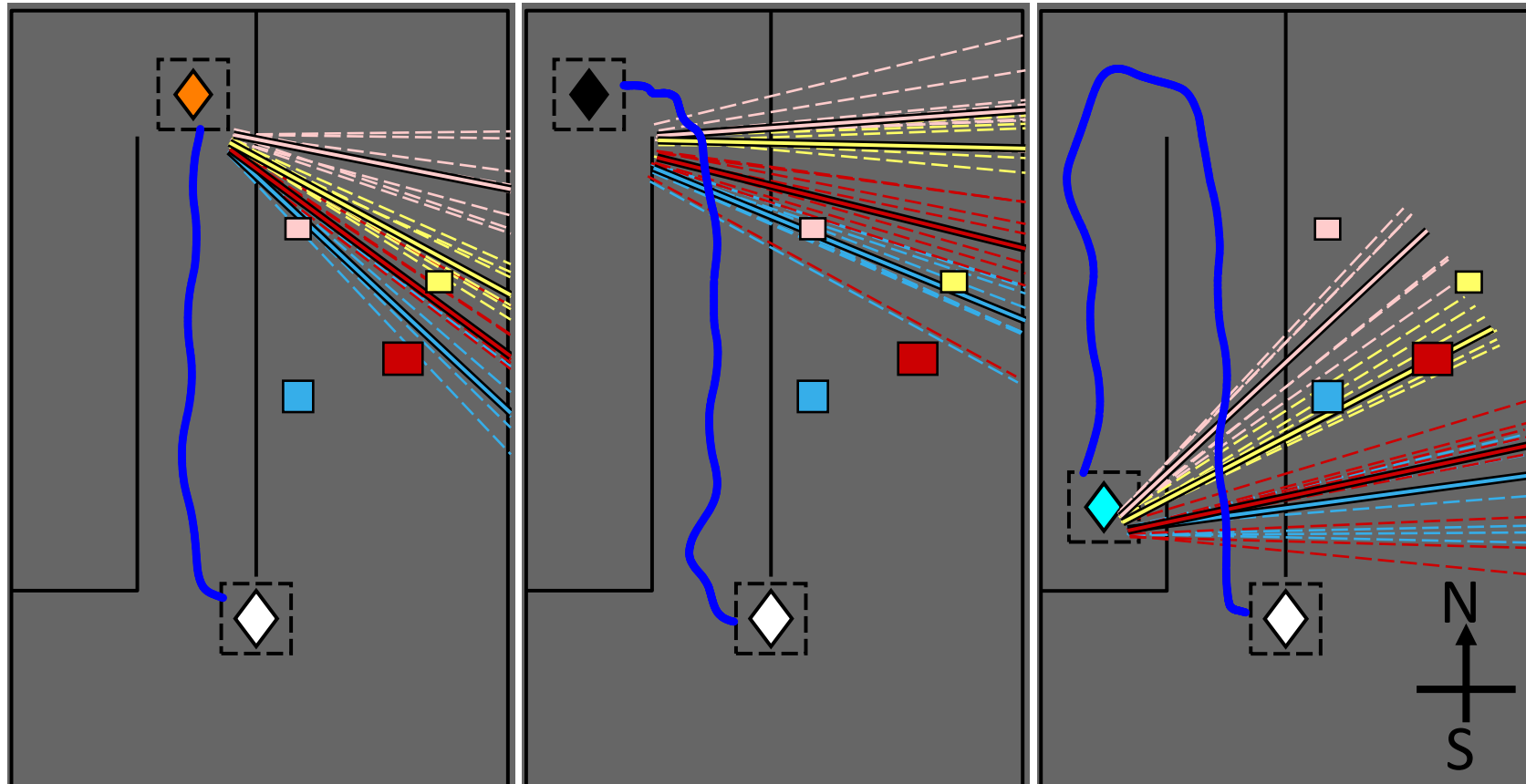


Jenny Vuong



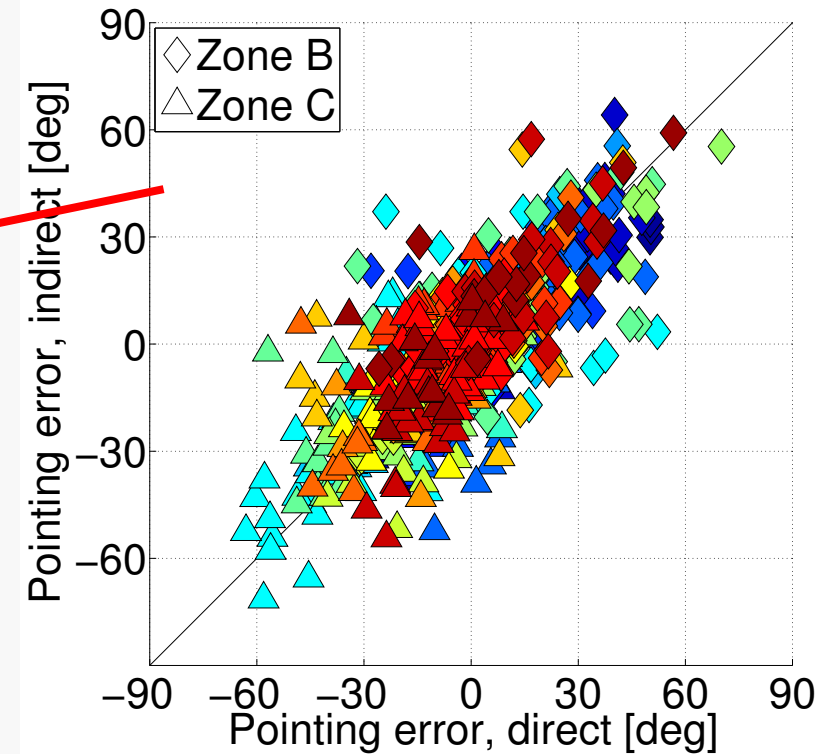
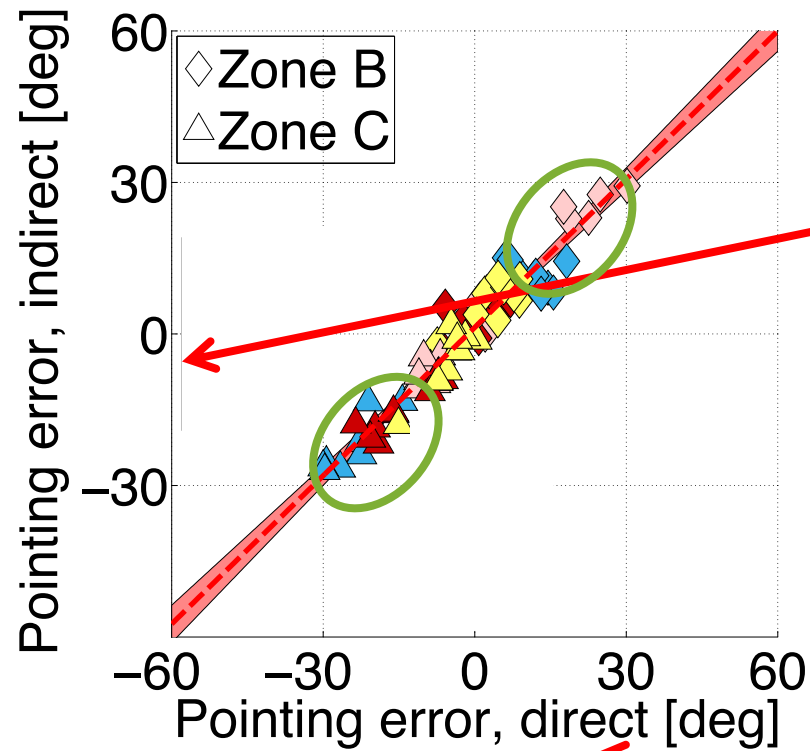
nVis SX111 HMD
Vicon tracking

People show large, consistent biases

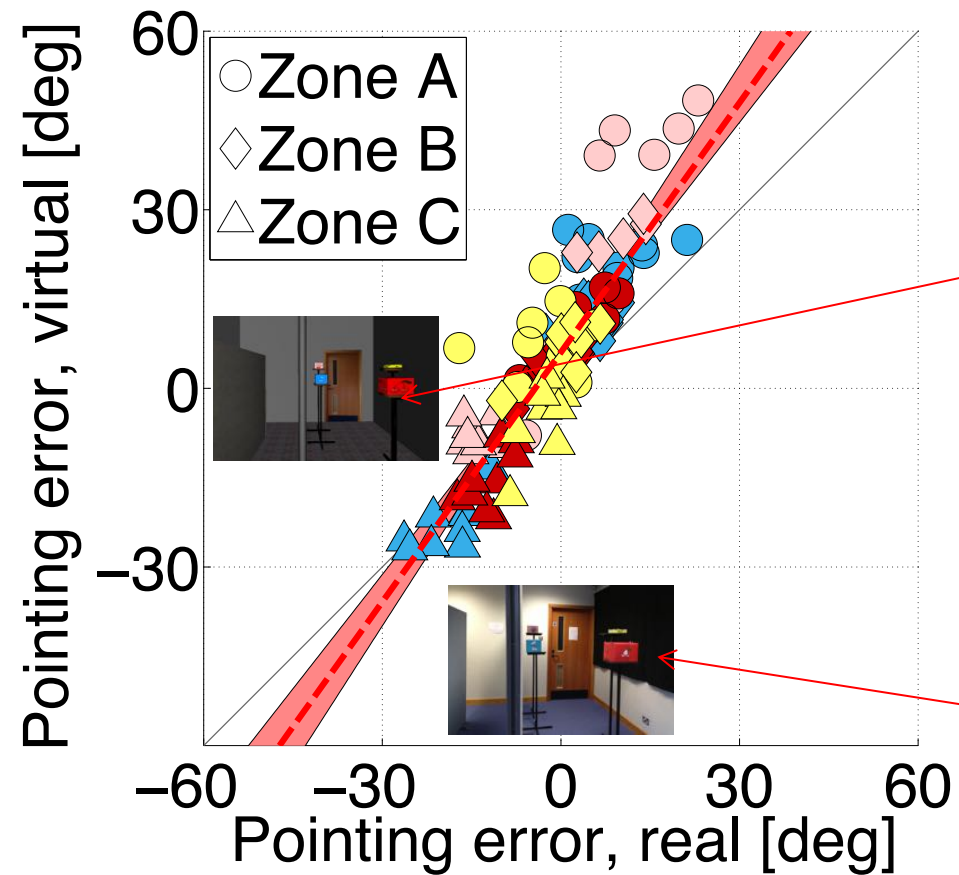


- Task:
 - view a scene
 - walk without any further view of the objects
 - point to the objects
 - easy to do if we update our location in a 3D reconstruction (SLAM)

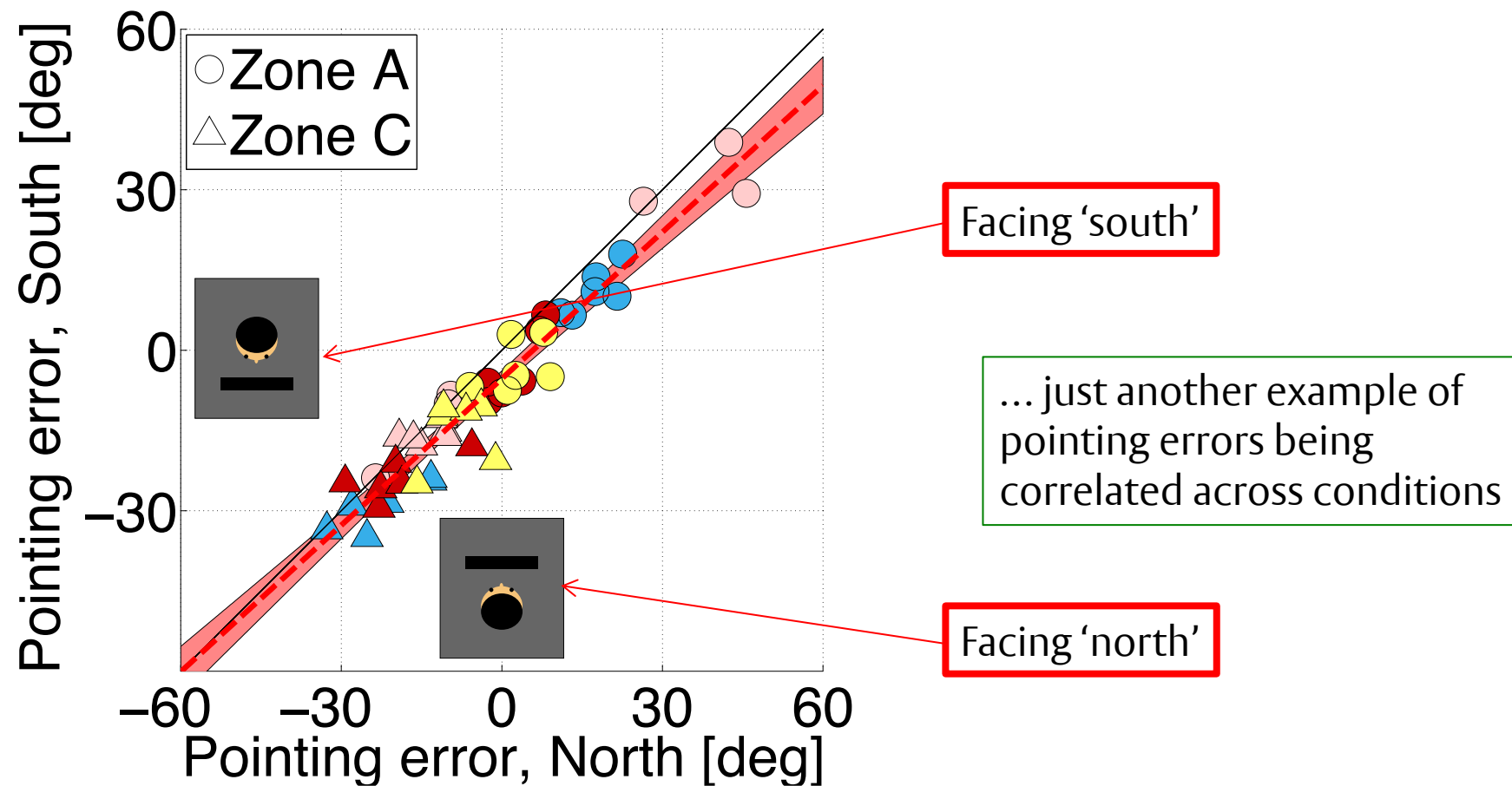
... independent of the route they take ...



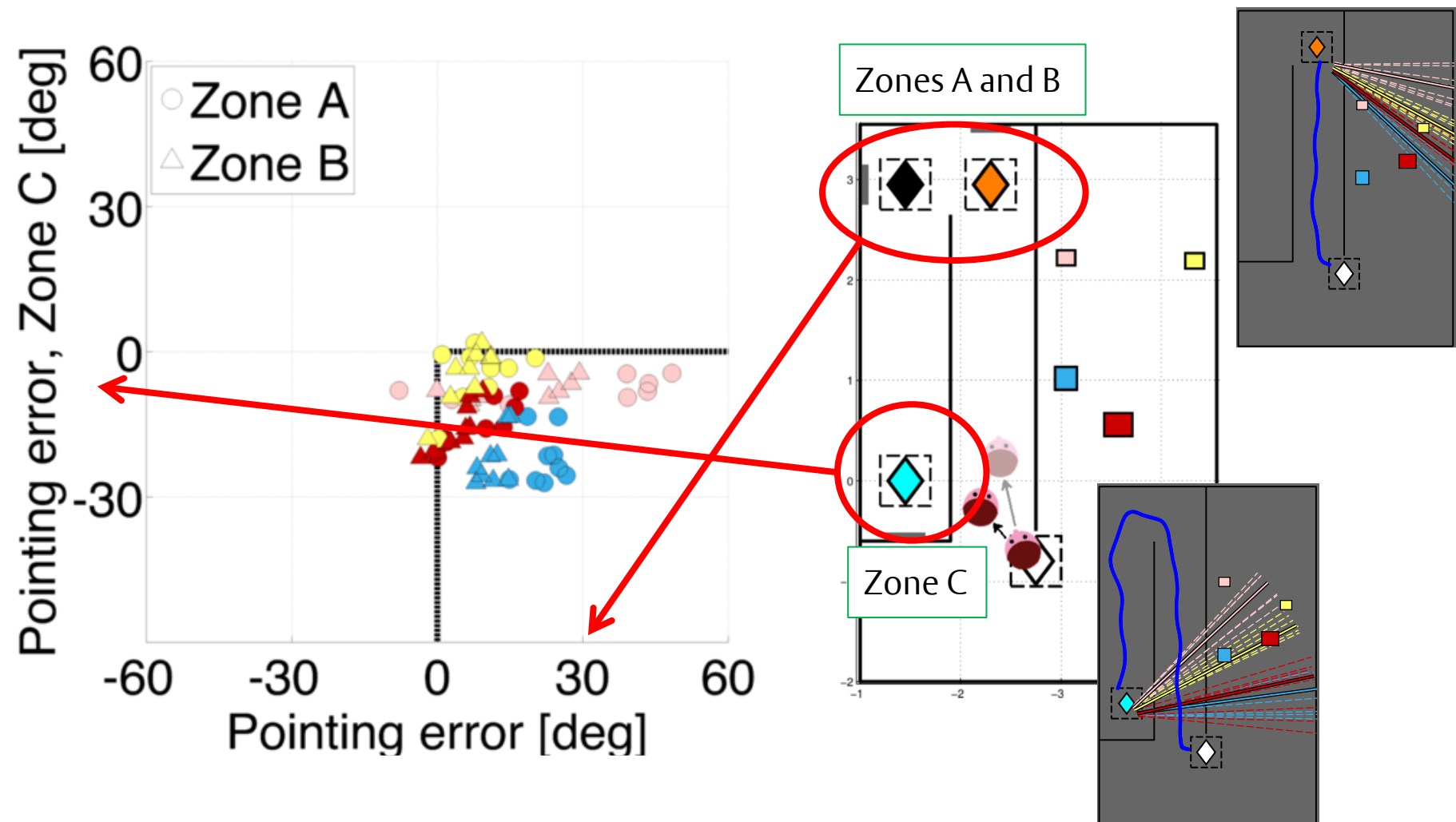
... similar biases in real and virtual worlds ...



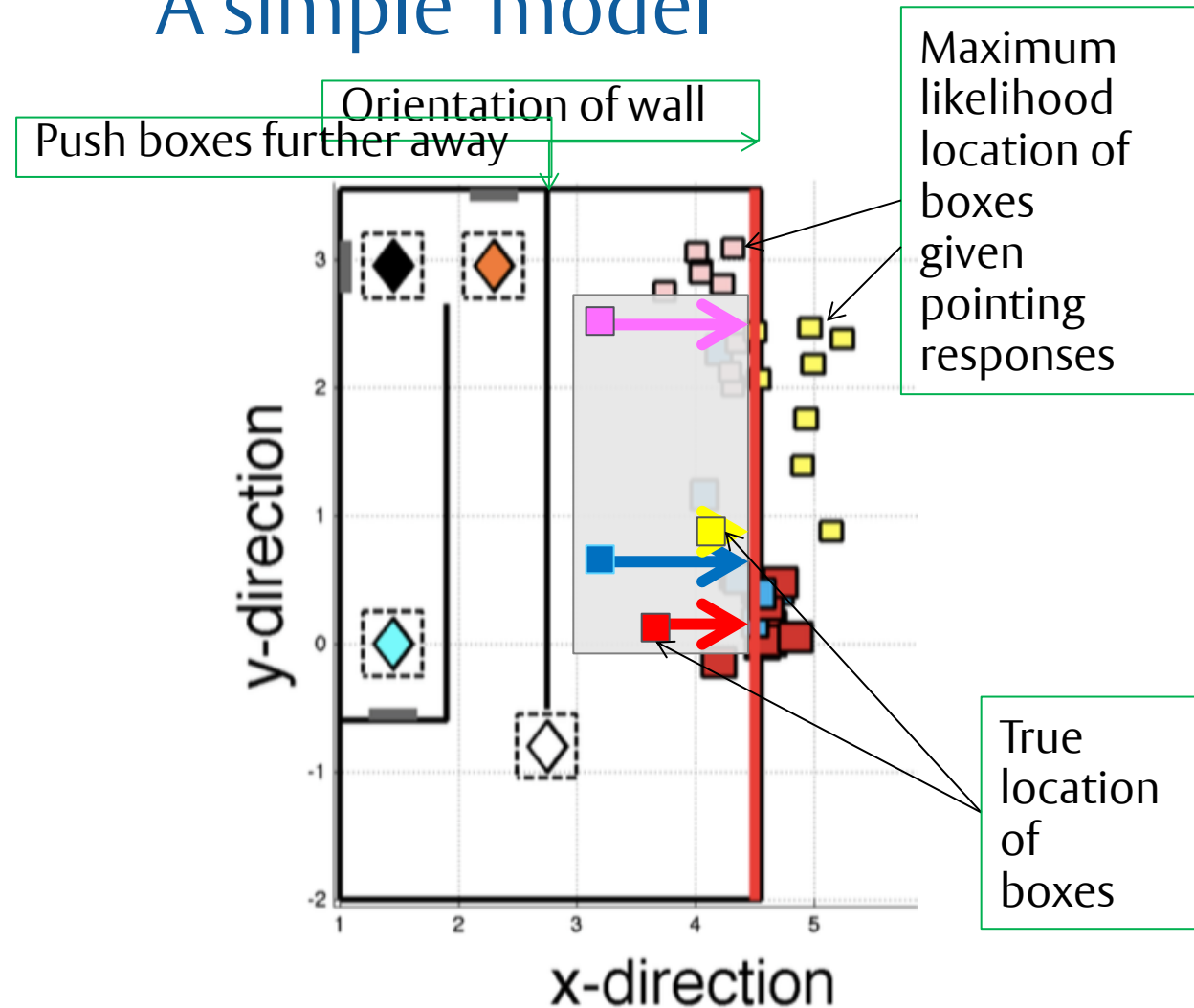
... whether looking 'north' or 'south' ...



... but heavily dependent on pointing zone

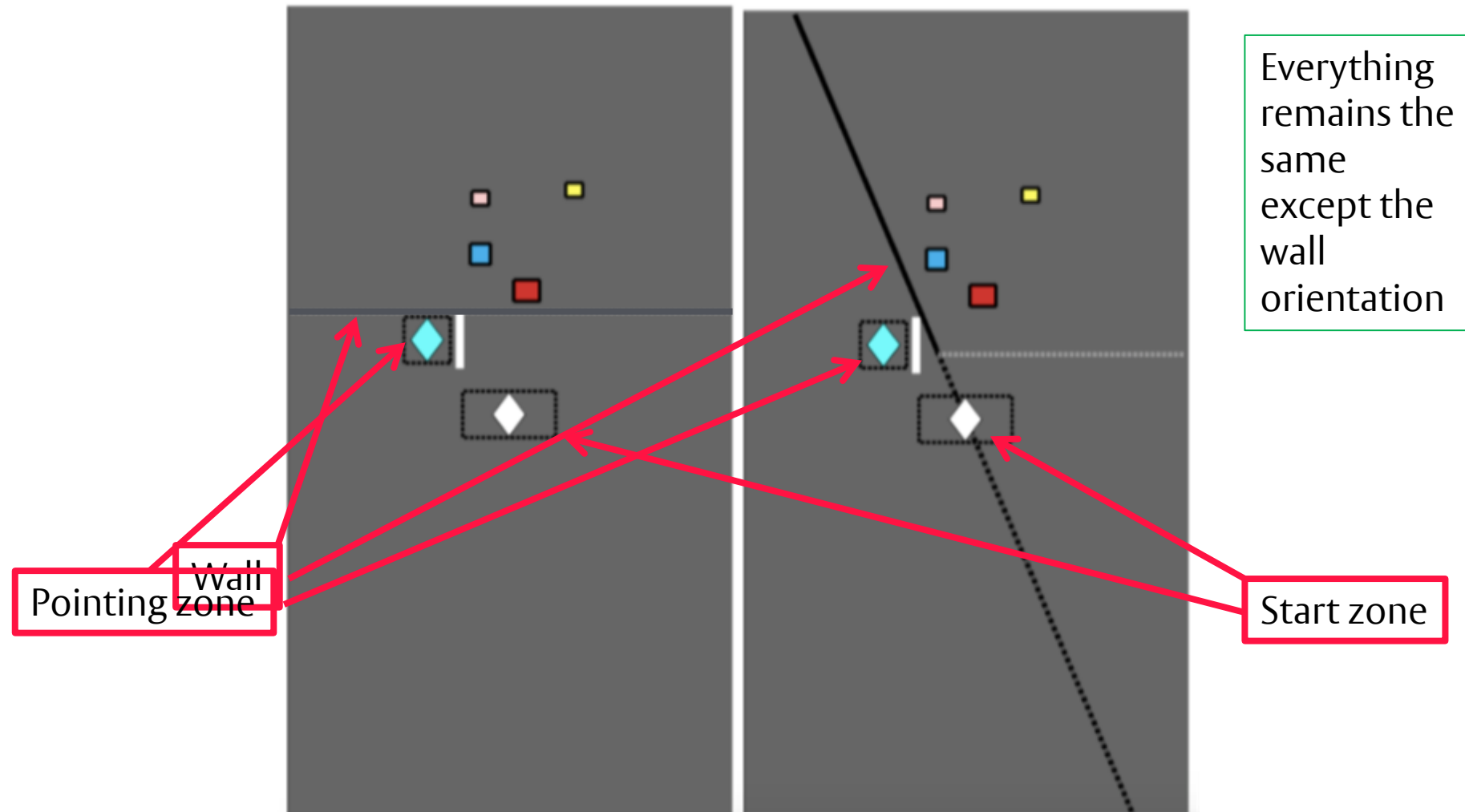


A simple 'model'

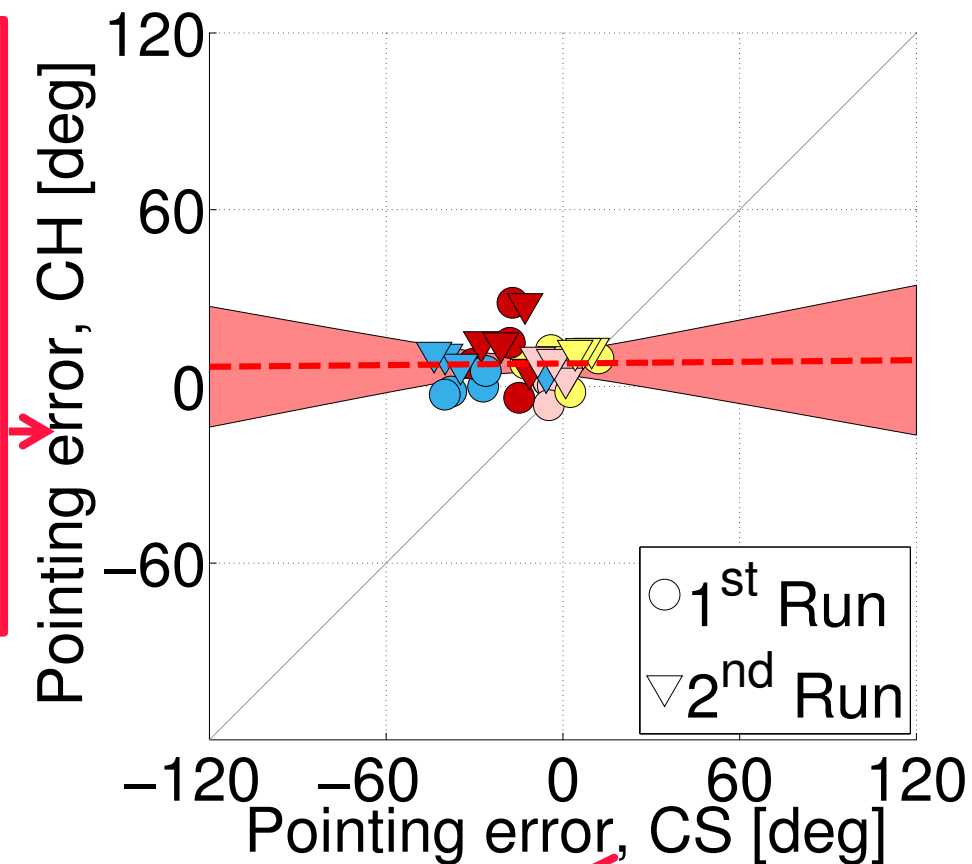
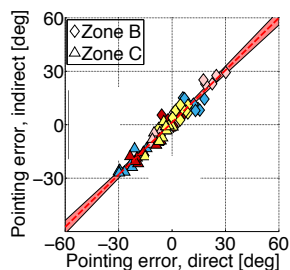


- Participants behave as if they ignore crucial aspects of the geometry of the scene
- pointing responses suggest they assume objects lie in a plane (or something close to this)

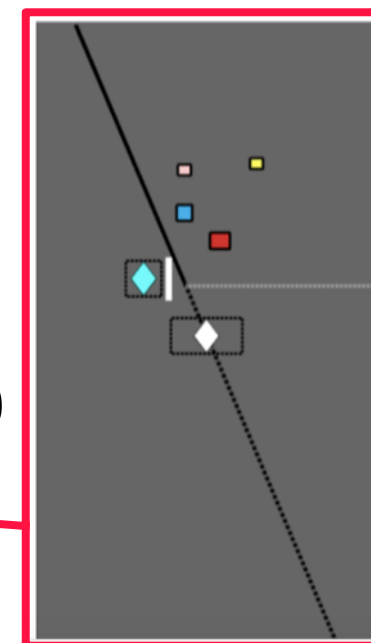
An effect of the wall orientation:



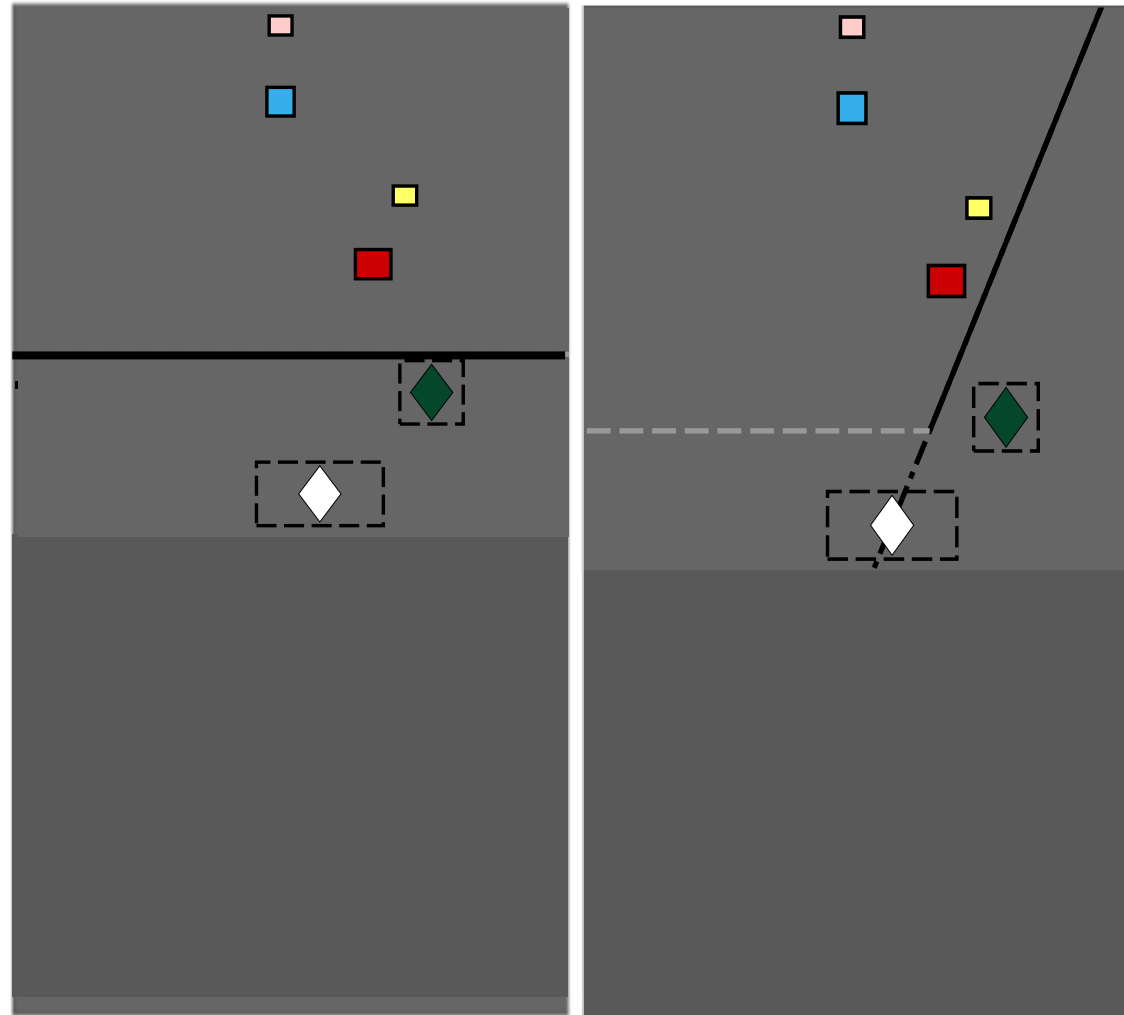
An effect of the wall orientation:



Everything remains the same except the wall orientation

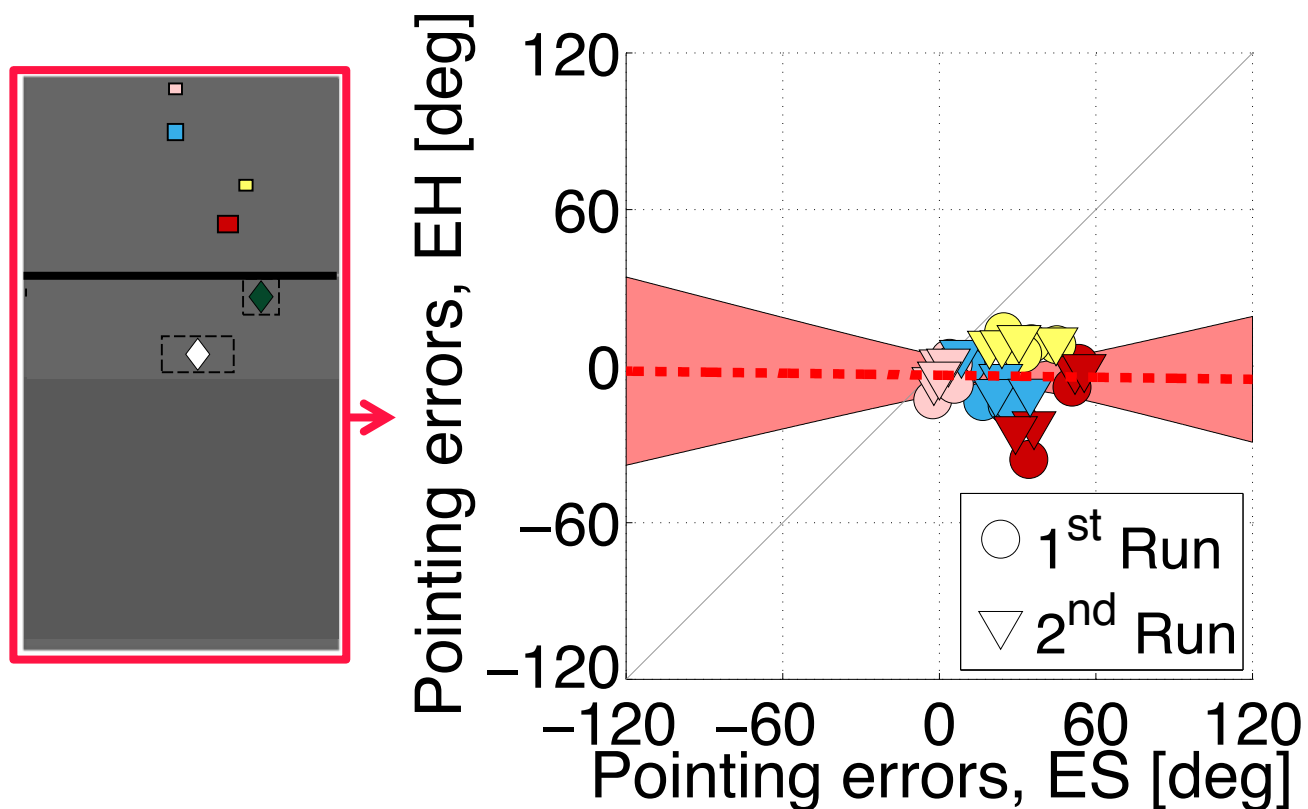


An effect of the wall orientation:

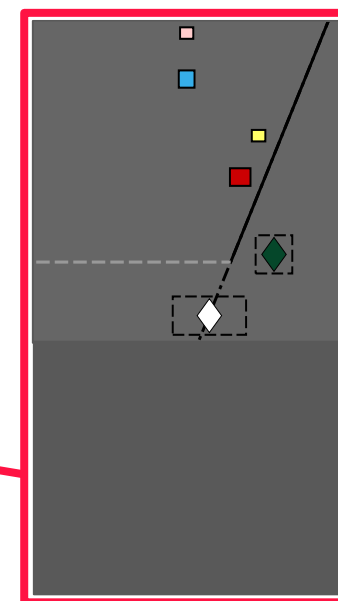


Also tested
other
viewing
zones,
other wall
orientations

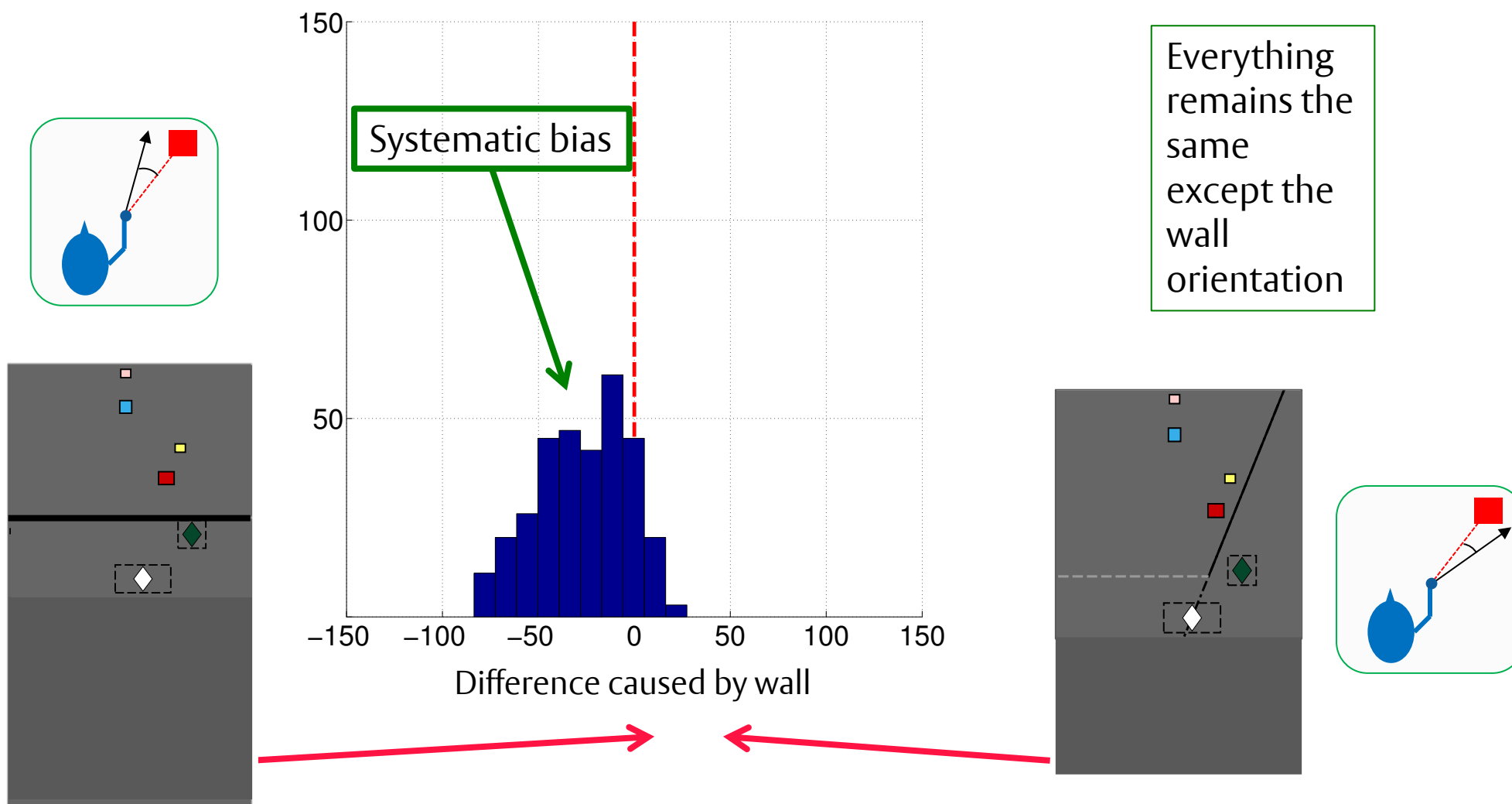
An effect of the wall orientation:



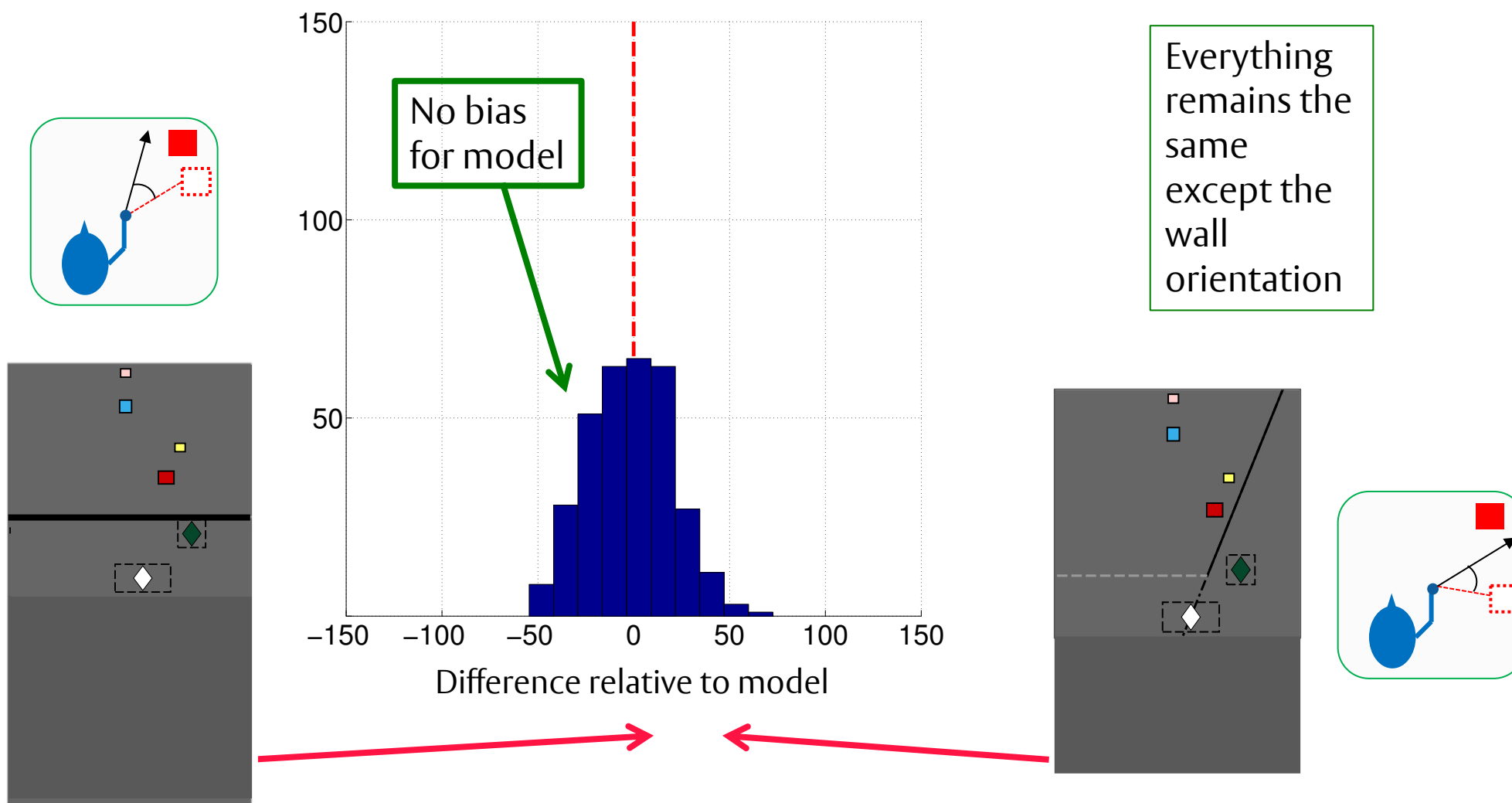
Everything remains the same except the wall orientation



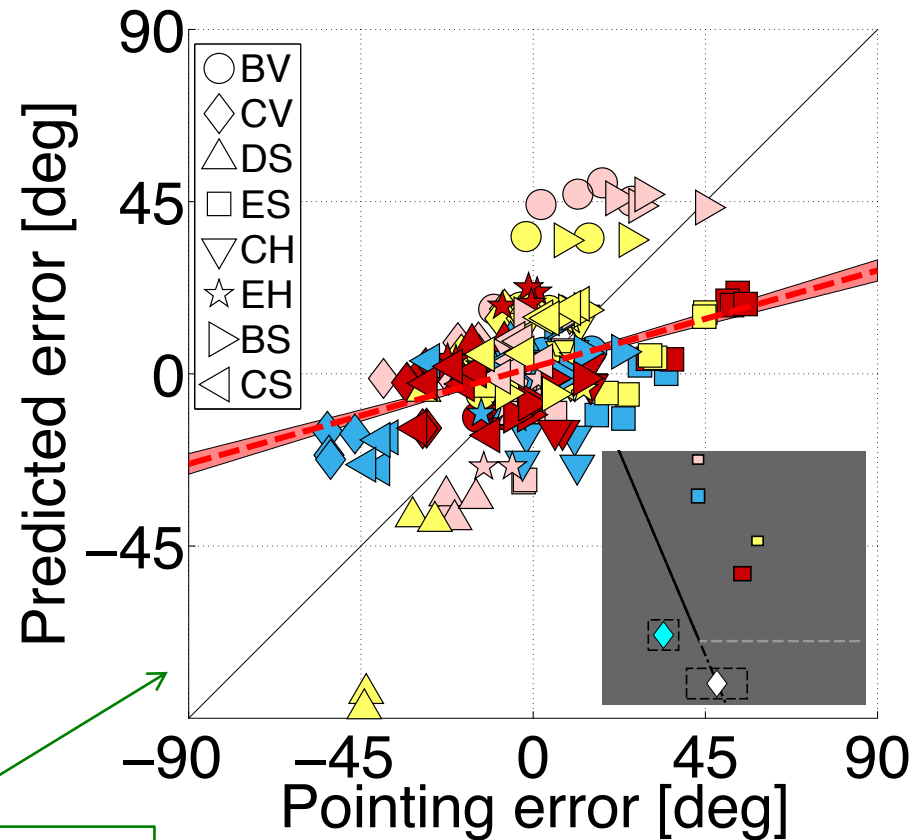
An effect of the wall orientation:



The model accounts for the effects of the wall



Give a 3D representation the best possible chance of explaining the data...



Predictions based on
maximum likelihood
location of boxes

Predictions based on
boxes-squashed-onto-
-a-plane model

Q: Can we update the visual direction of
unseen objects as we move?

A: not very well (we have poor heuristics
for imagining)



Jenny Vuong



‘Neural rendering’ without a 3D reconstruction

Neural Scene Representation and Rendering

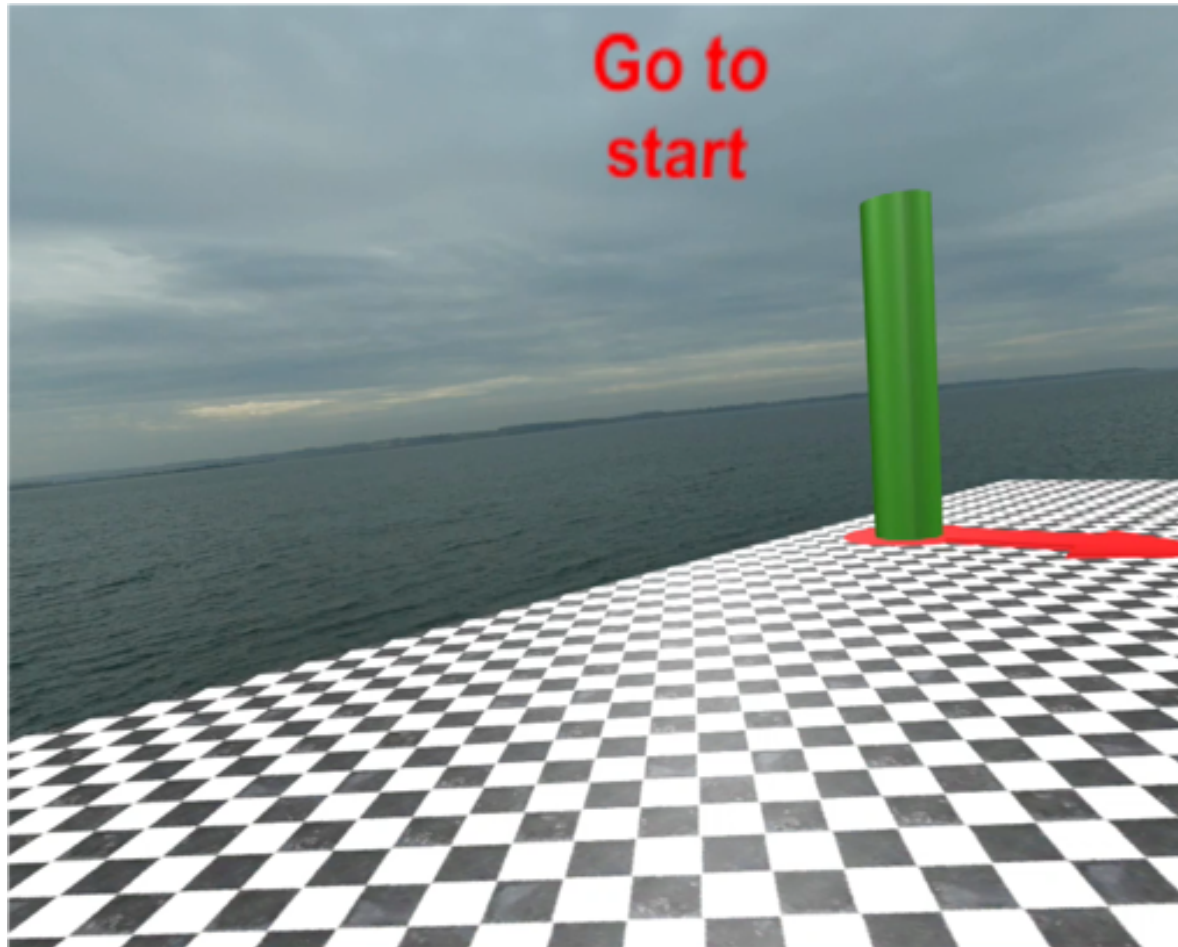
S. M. Ali Eslami*, Danilo J. Rezende*, Frederic Besse, Fabio Viola, Ari S. Morcos, Marta Garnelo, Avraham Ruderman, Andrei A. Rusu, Ivo Danihelka, Karol Gregor, David P. Reichert, Lars Buesing, Theophane Weber, Oriol Vinyals, Dan Rosenbaum, Neil Rabinowitz, Helen King, Chloe Hillier, Matt Botvinick, Daan Wierstra, Koray Kavukcuoglu and Demis Hassabis





Alex Murry

Learning to point to targets in a maze

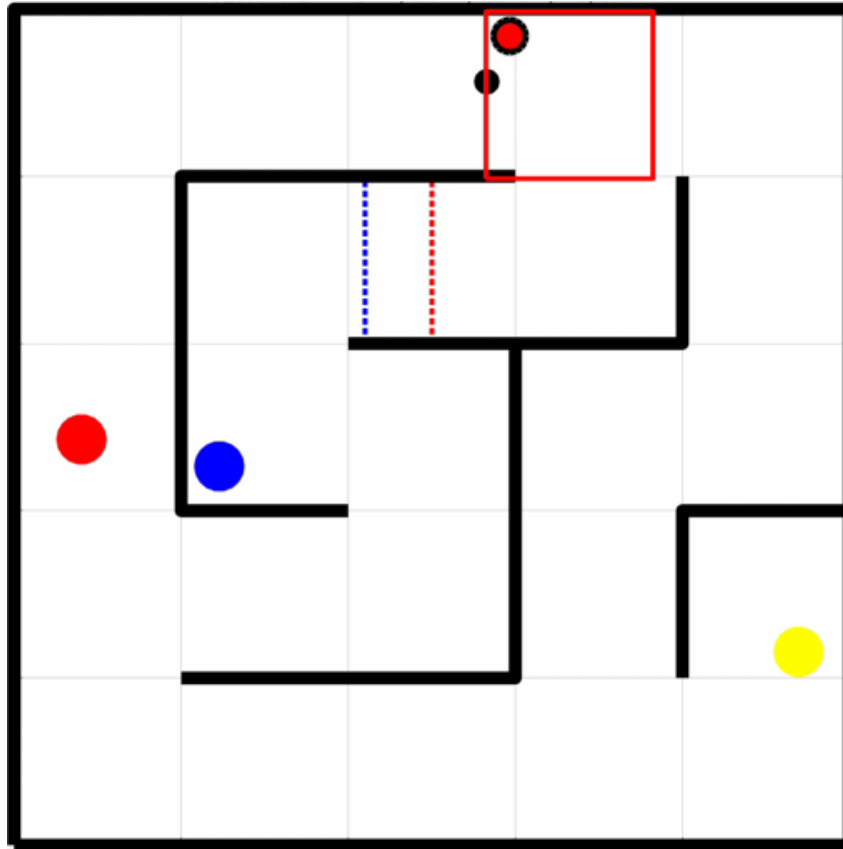


- Tasks:
- (i) find targets in specified order and
 - (ii) point to them...



Alex Murry

Learning to point to targets in a maze



Life gets harder...

Learning phase (repeat x5):

- a) Navigation: go Start-R-G-B-Y
- b) Pointing: from Y point to S, R, G, B

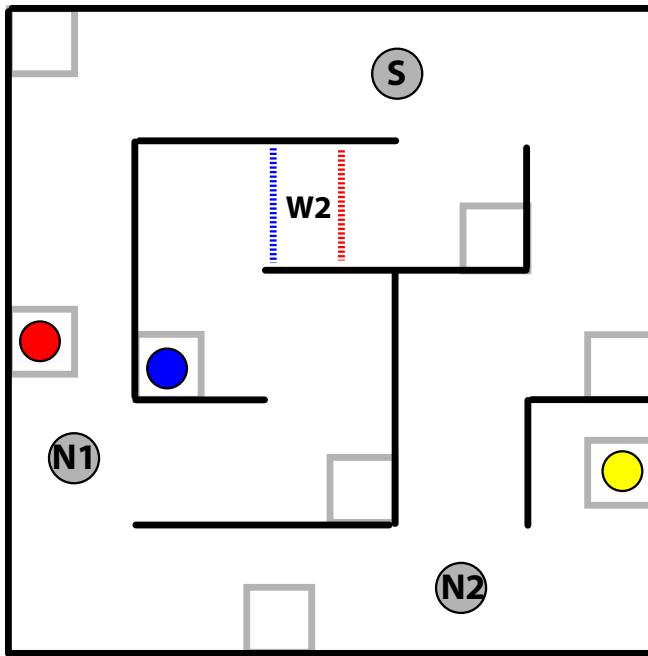
Test phase (x3):


- a) Random sequences
- b) Point to all targets



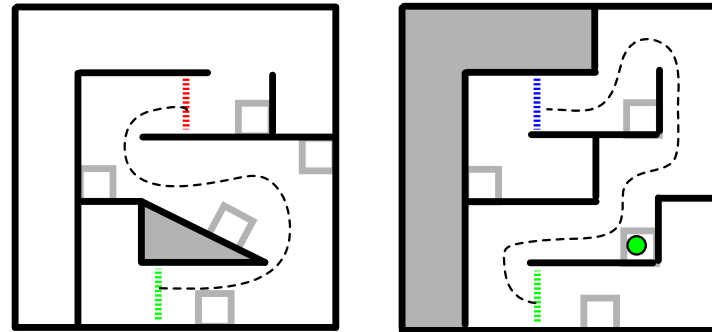
Alex Murry

Non-metric scene: 1 wormhole

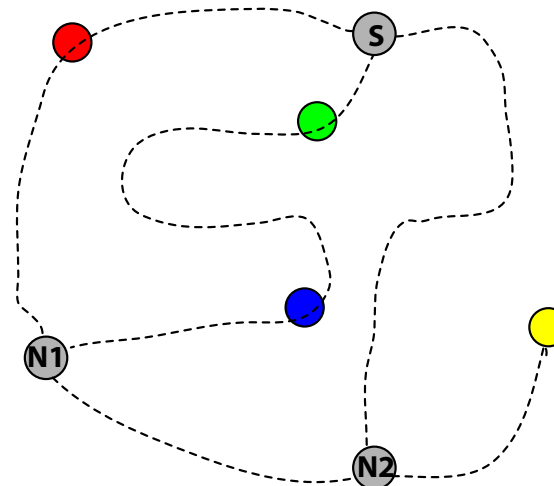


- - topological node
- - small walls
-  - wormhole

wormhole

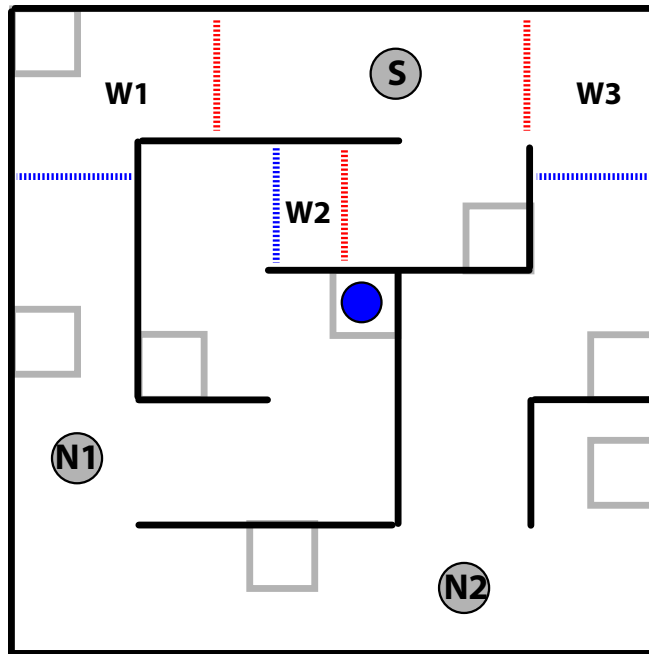


topological graph

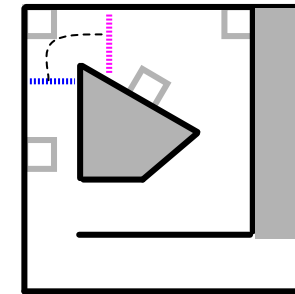
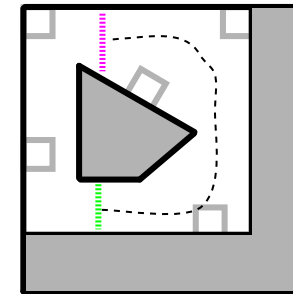
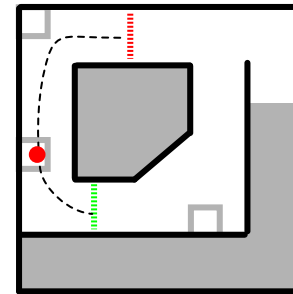




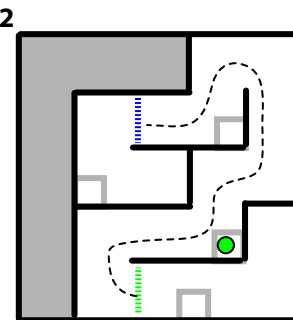
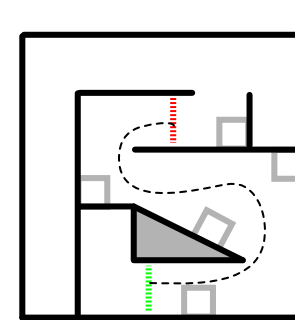
Non-metric scene: 3 wormholes



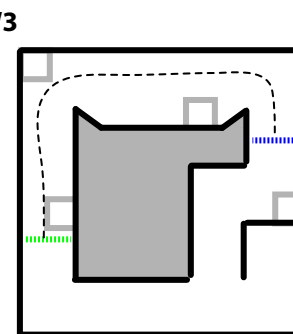
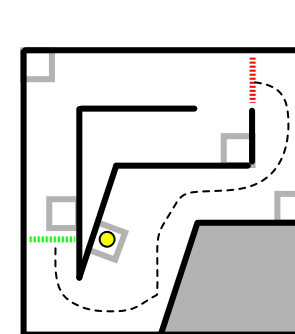
- - topological node
- - small walls
- W - wormhole



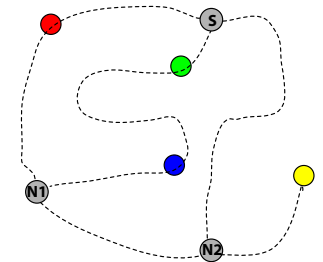
Wormhole 1



Wormhole 2

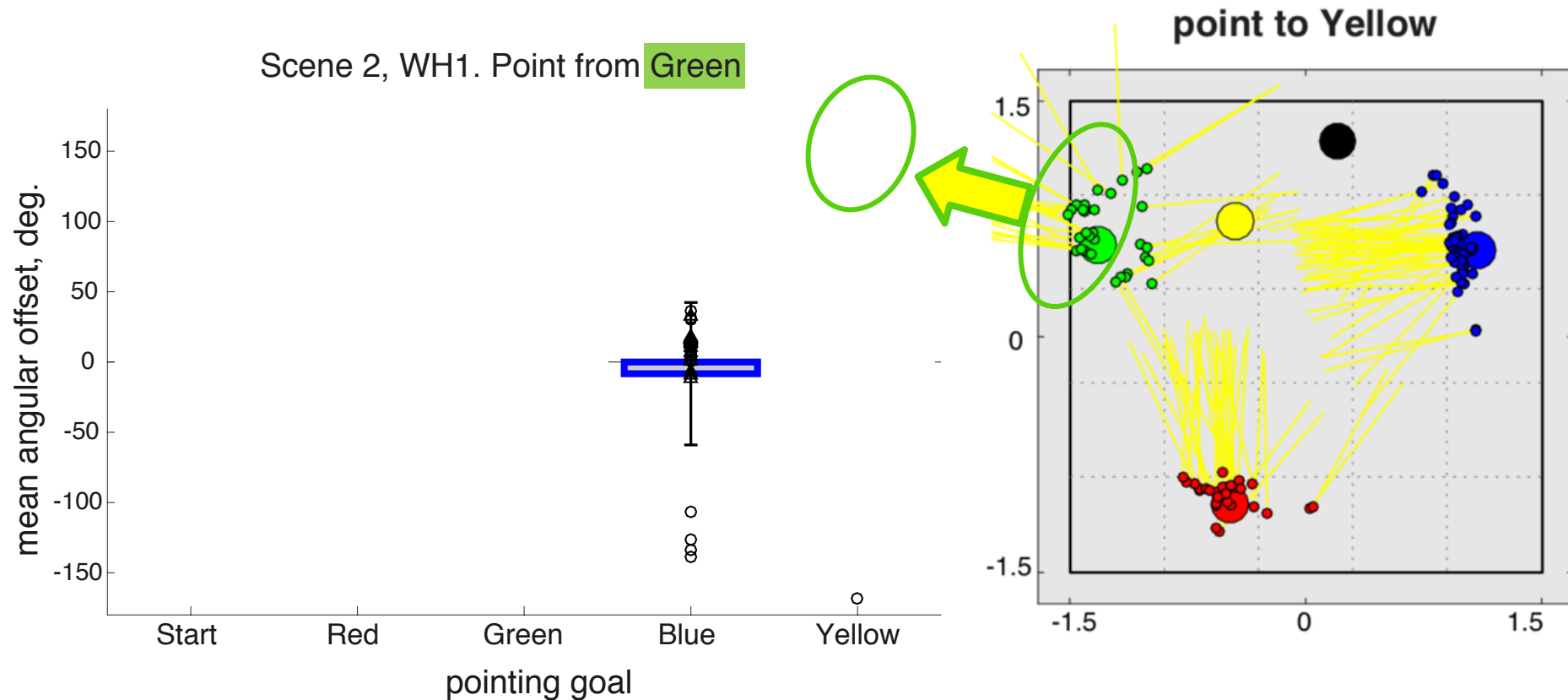


Wormhole 3



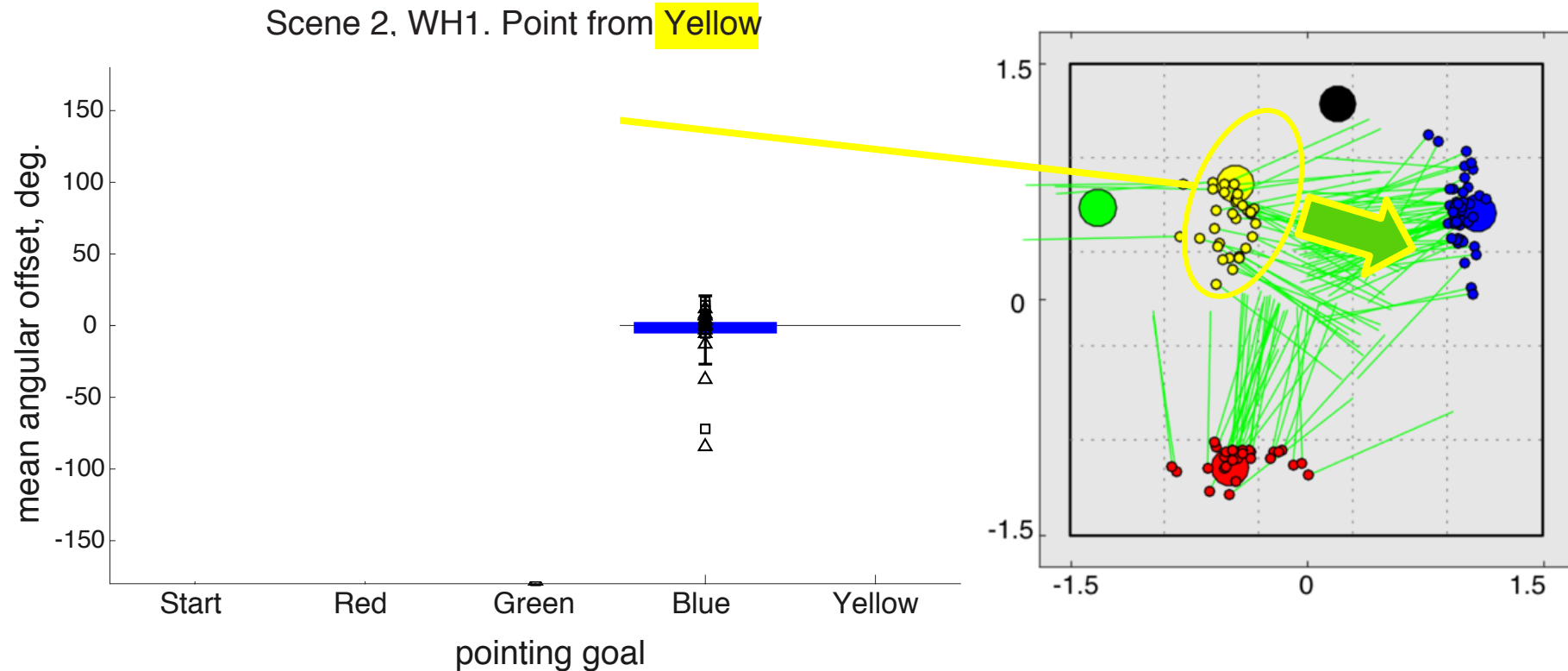
... but same
topological
structure

Pointing: a 'metric' task



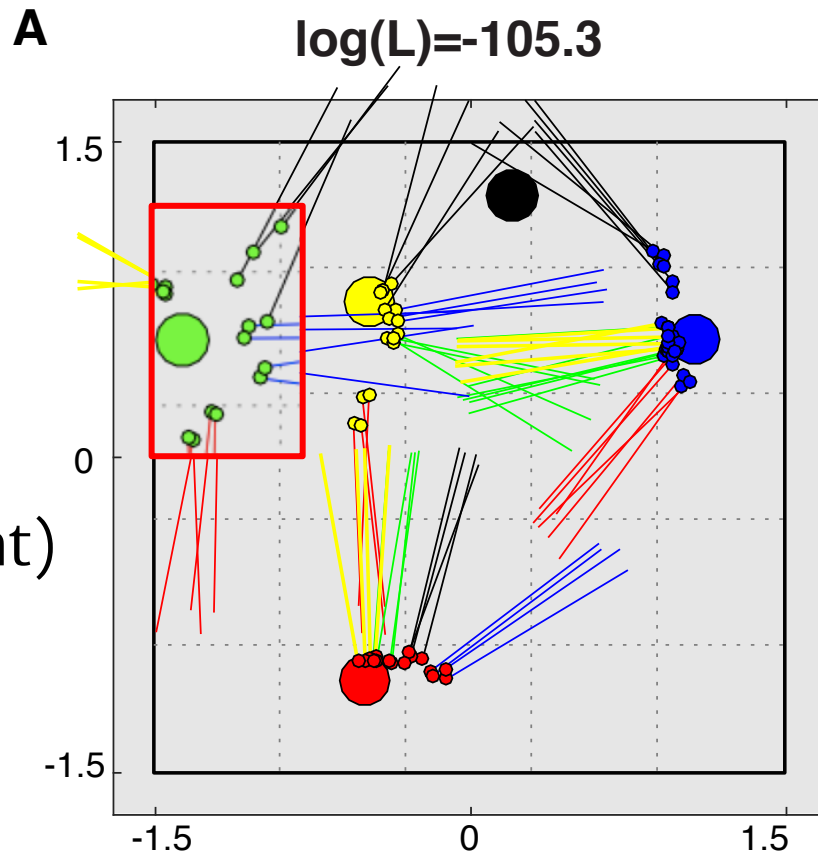
Pointing to some targets leads to very large, systematic errors.

Pointing: a 'metric' task



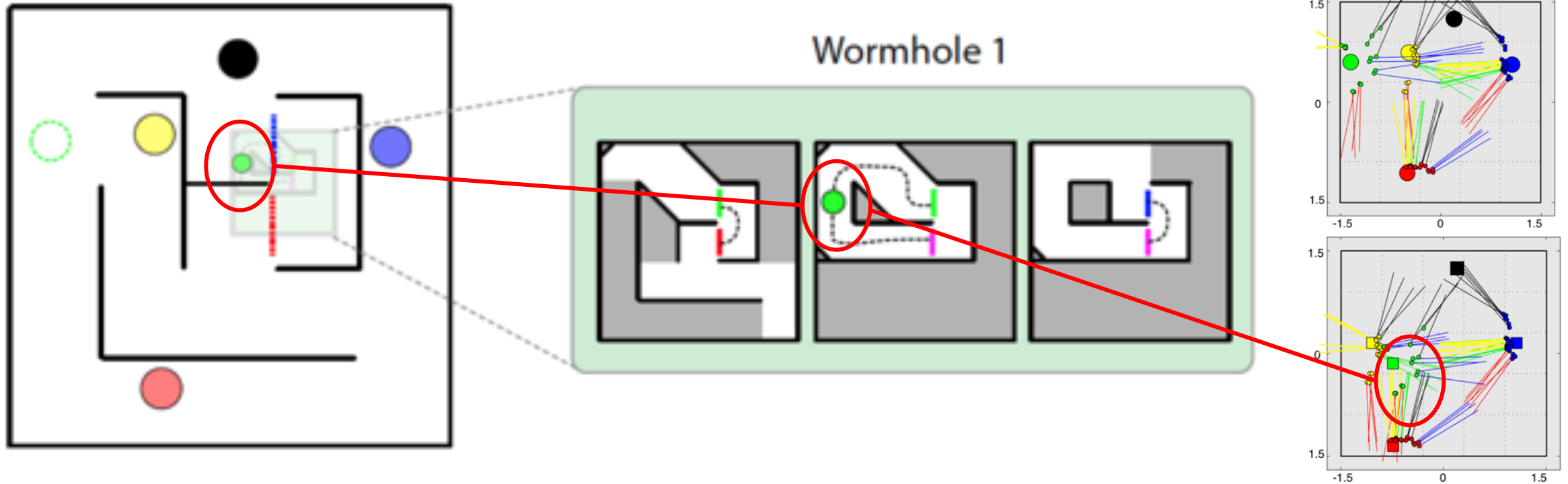
Pointing to some targets leads to very large, systematic errors.

Pointing: a 'metric' task



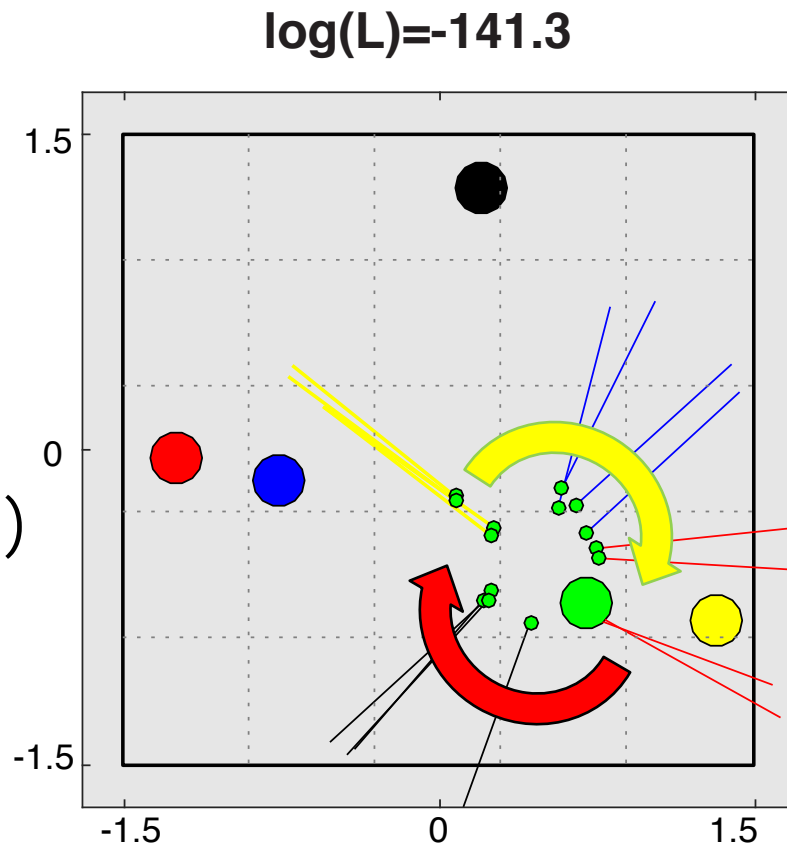
In the most likely configurations, green is to the east of yellow.

Pointing: a 'metric' task



It seems as if participants 'squash' the wormhole corridors into a smaller region than they actually occupy .

Adding in rotation



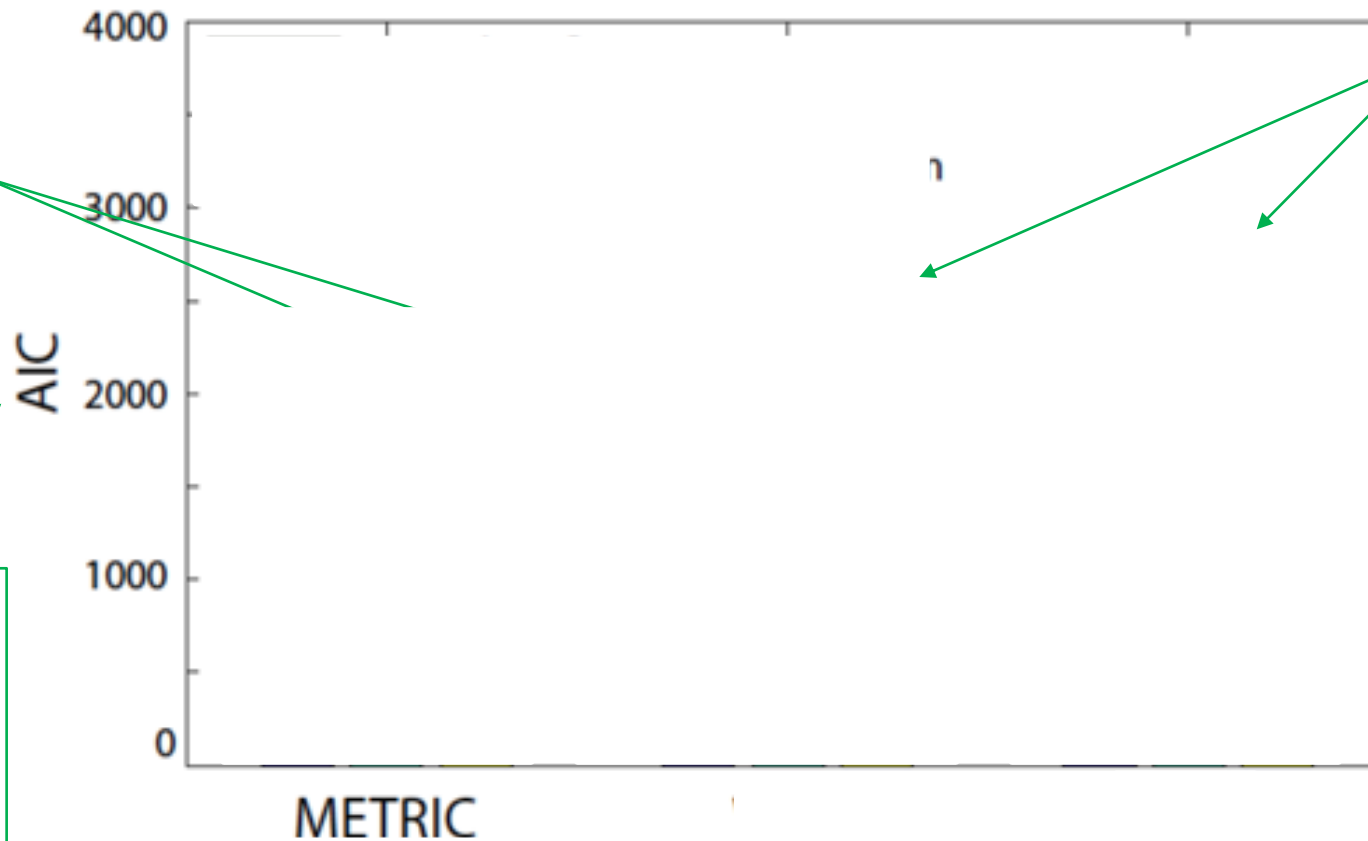
180° rotation
of all pointing
directions

This takes into account the possibility that people are disoriented.
But it is not compatible with a single, consistent 3D representation.

Model comparison

In the METRIC condition, optimised translation and rotation are not better models than the original configuration (when penalized for the extra parameters in the models)

Akaike information criterion, a measure of likelihood (low is more likely)



In the WORMHOLE conditions, the best model is one that optimizes the location of the targets (i.e. a distorted world) *and* optimizes the rotation of the observer independently at each pointing zone

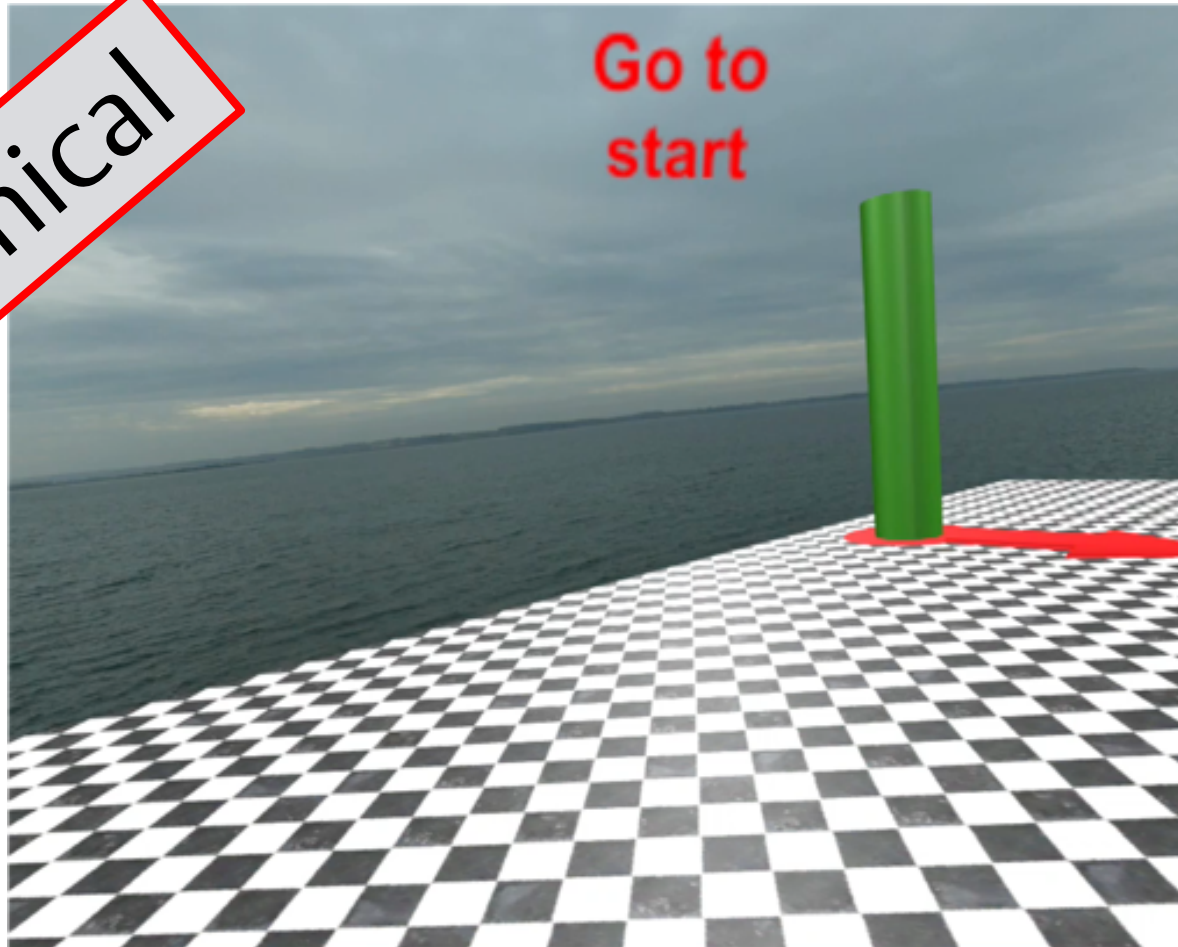
This is not a consistent, metric, 3D representation



Alex Murry

Learning to point to targets in a maze

Hierarchical

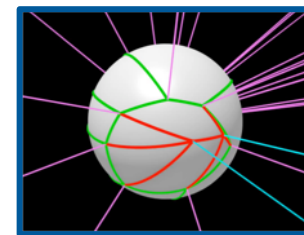


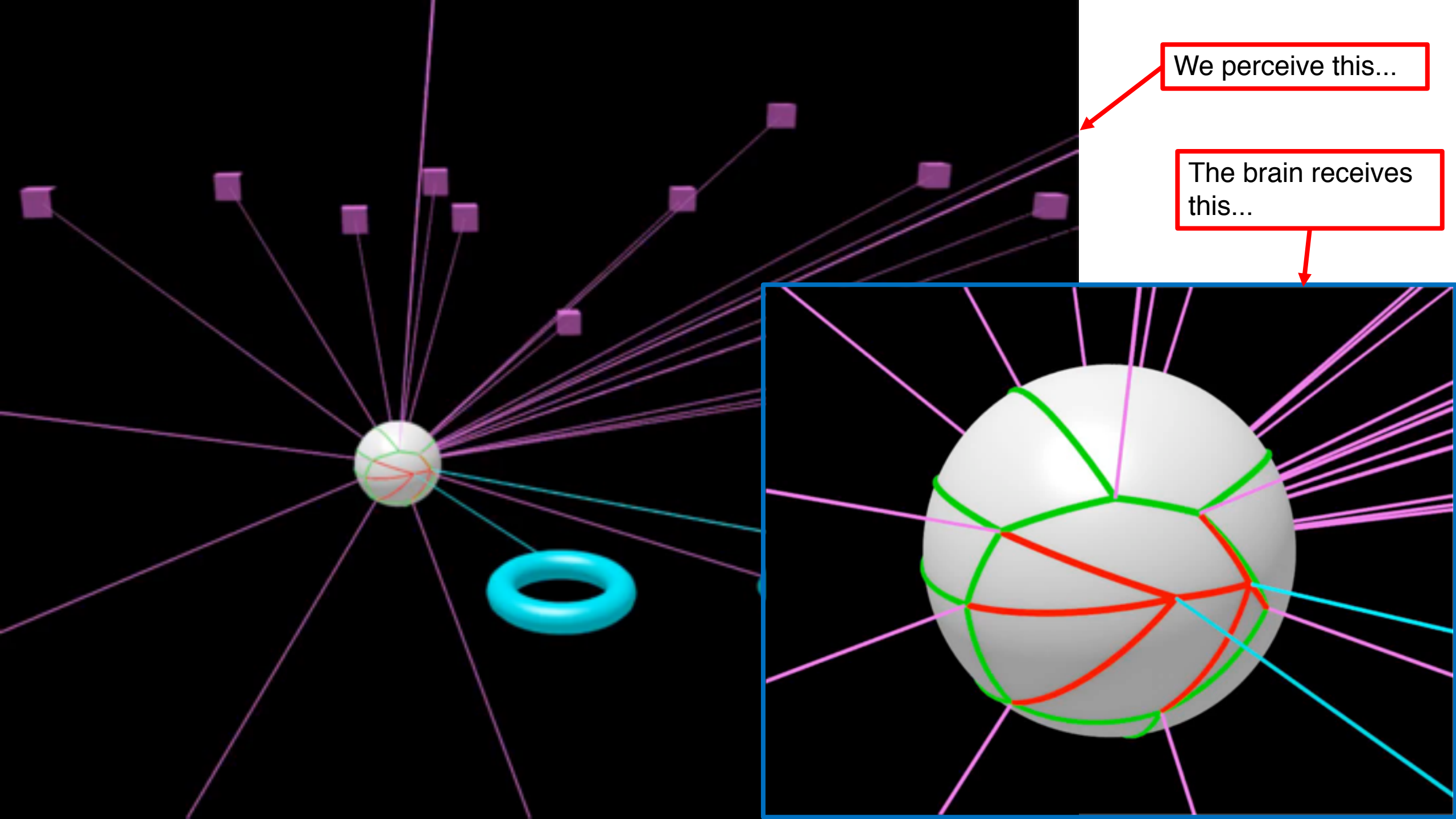
Go to
start

People's ability to point at unseen targets may be built up from an initial topological representation with information about lengths and turns gradually added as they learn about the environment.

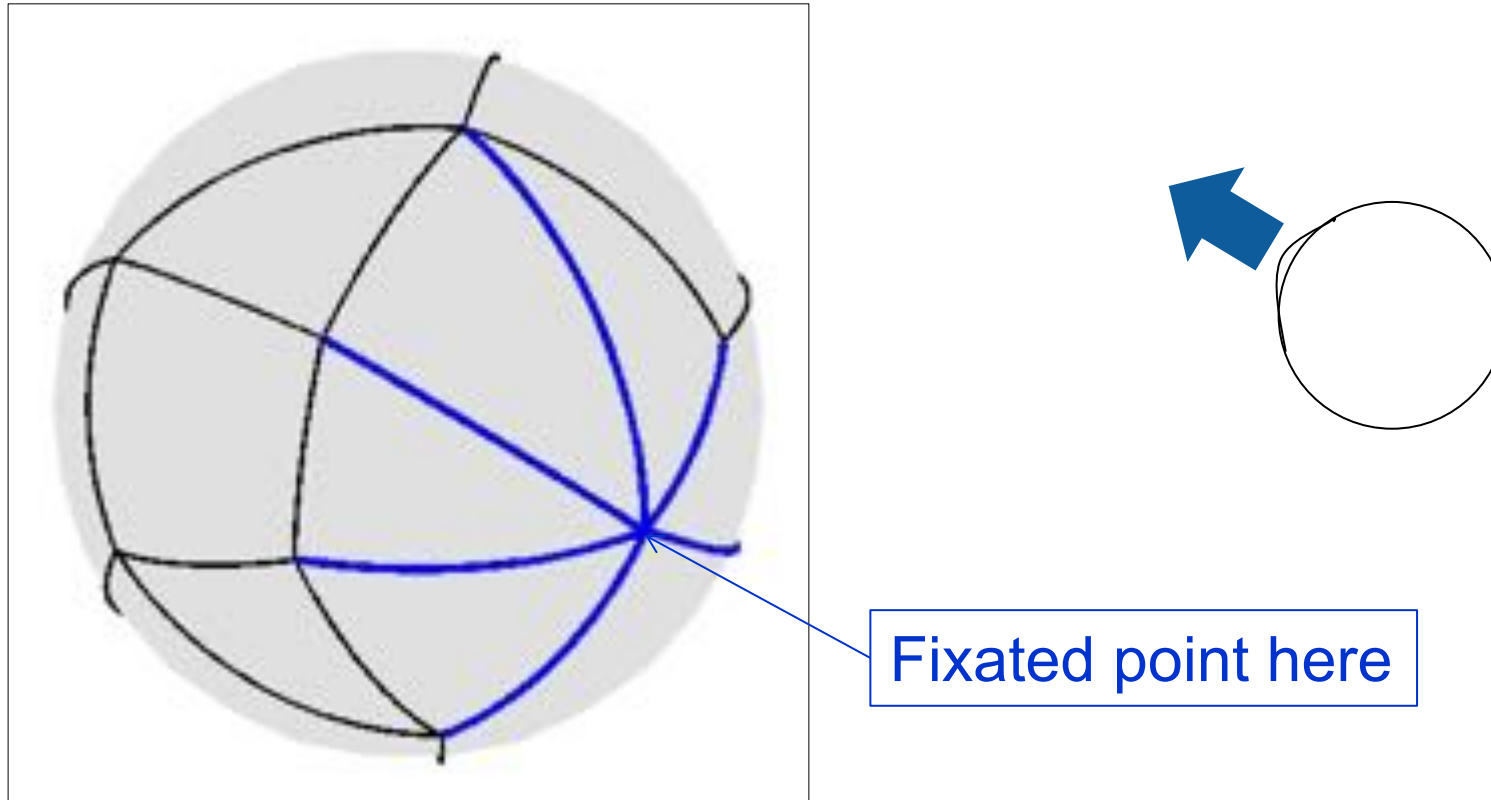
Outline

- Updating visual direction
 - some evidence and a ‘model’
- Navigating through wormholes
 - a 3D model is not the best explanation
 - coarse to fine learning of space
- A sphere of visual directions
 - information about viewing distance
 - A 2½ -D sketch

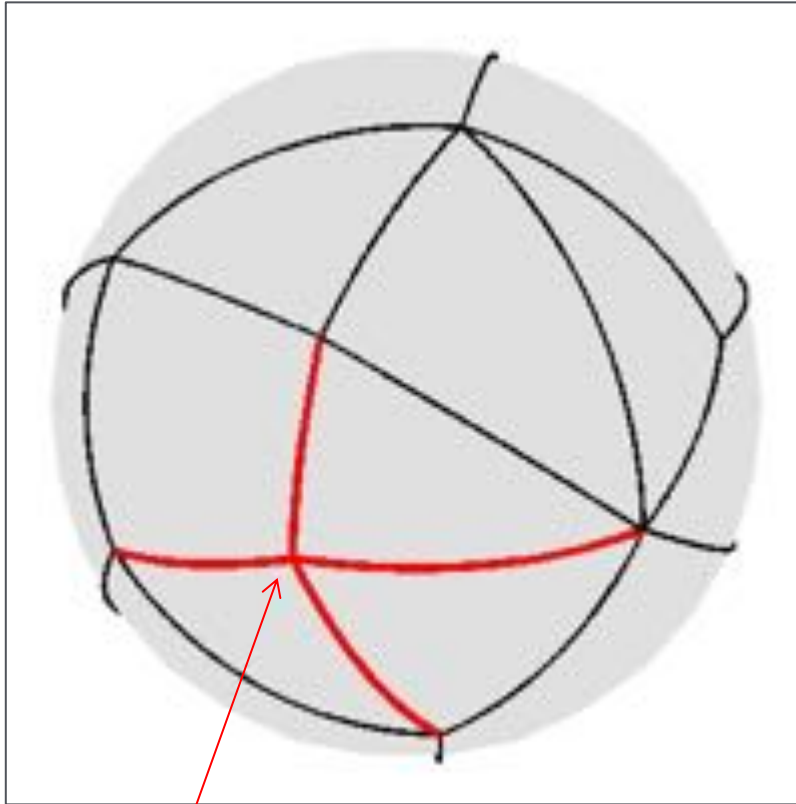




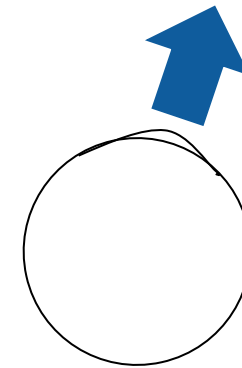
A stable coordinate frame for eye rotation



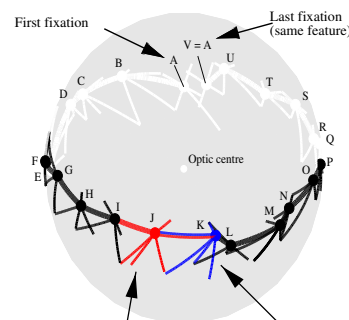
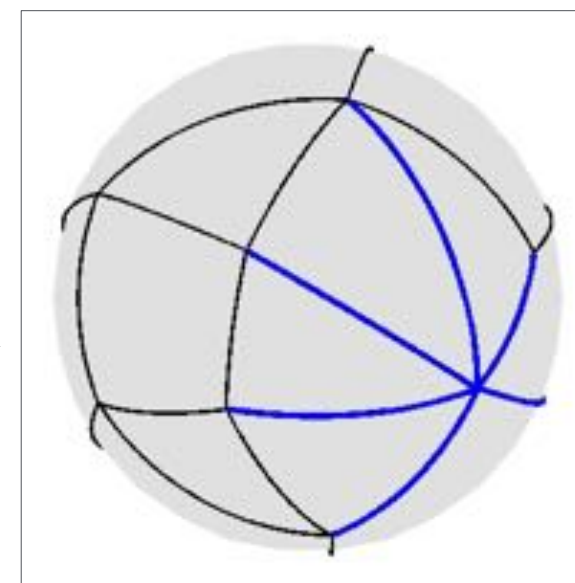
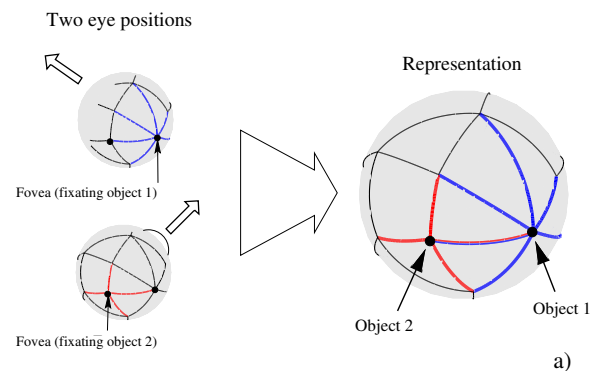
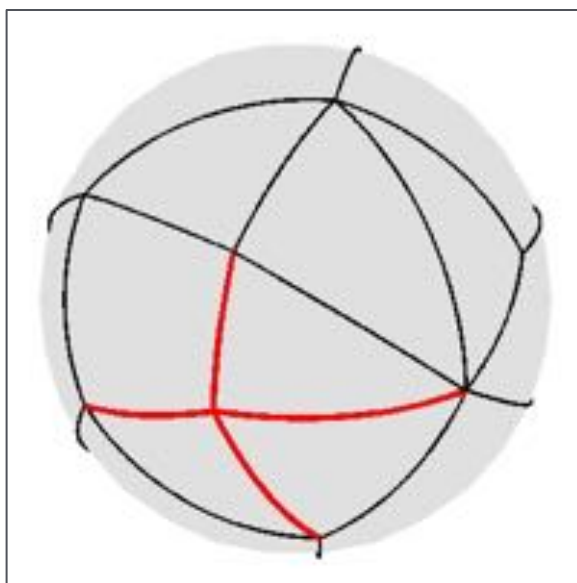
A stable coordinate frame for eye rotation



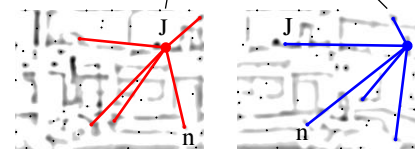
Fixated point here



A stable coordinate frame for eye rotation

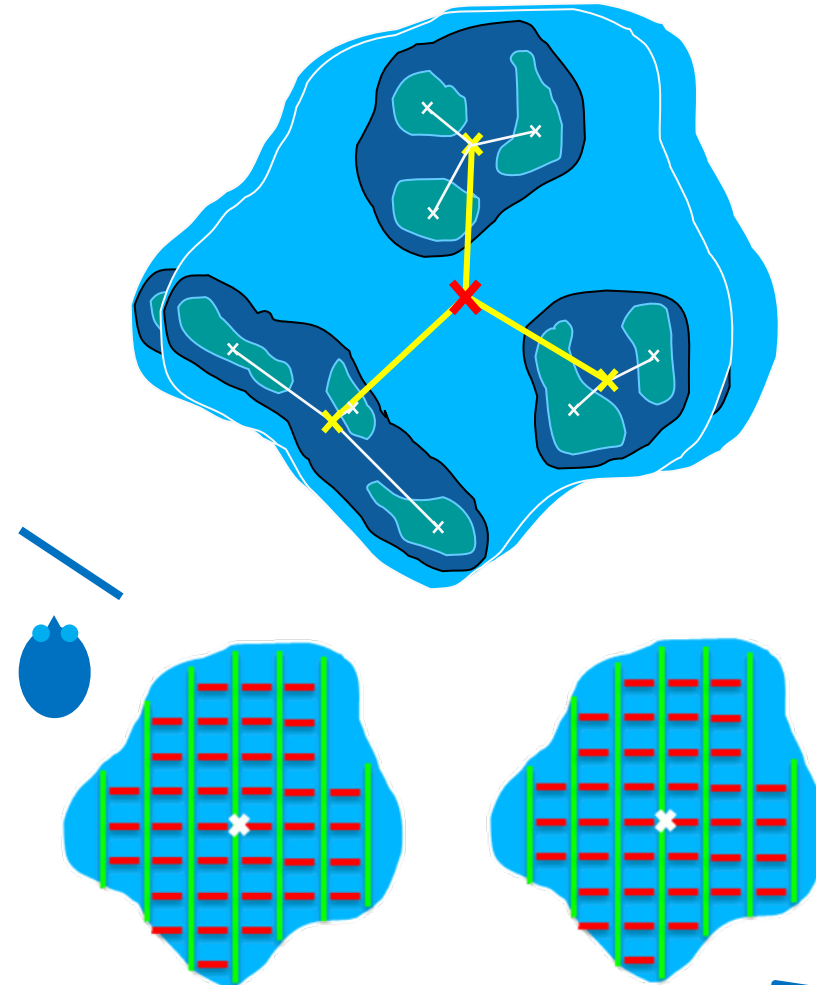
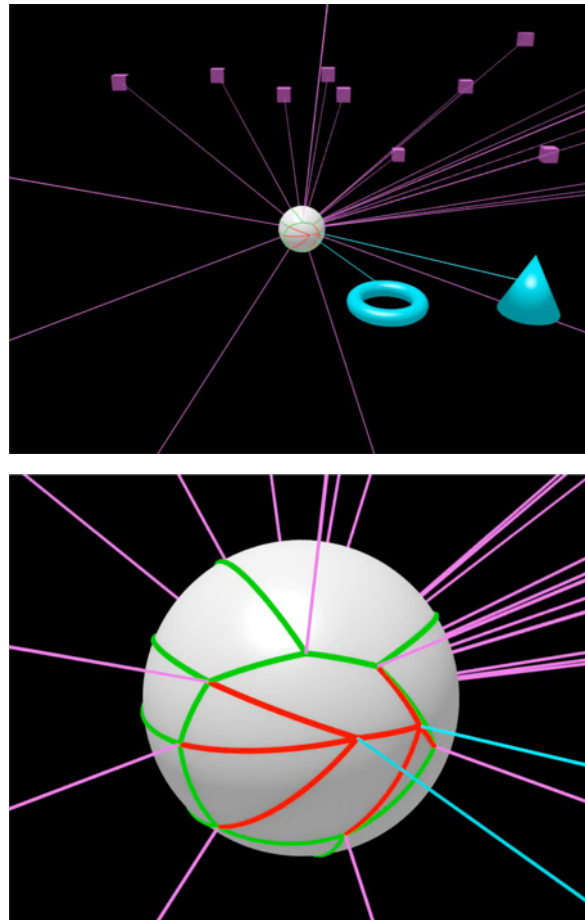


Glennerster, Hansard and Fitzgibbon (2001)

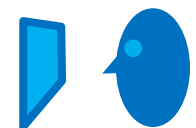


b)

Elasticity – a property that persists



Not just another description of optic flow. Instead, it is a long-lasting useful description with predictive power

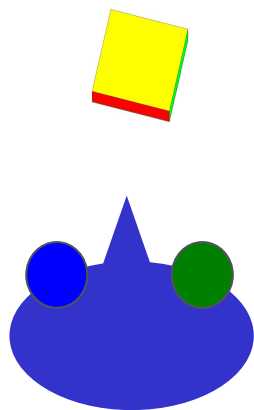
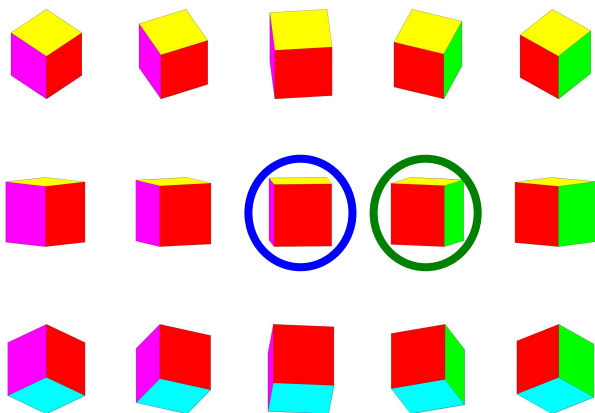




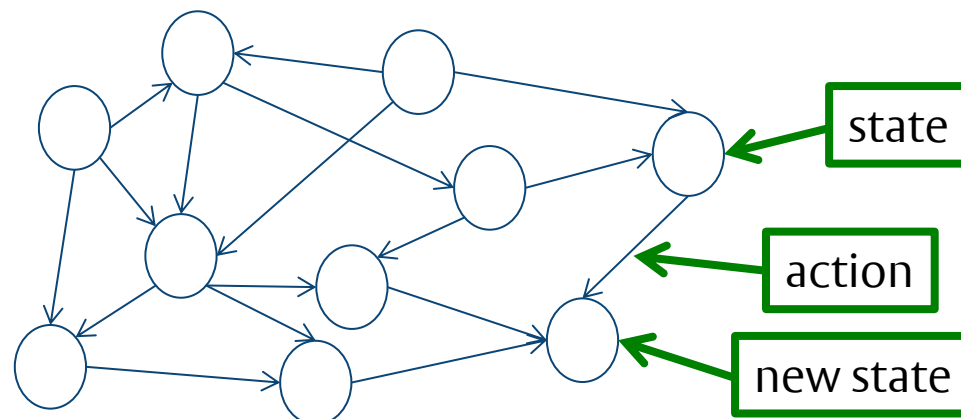
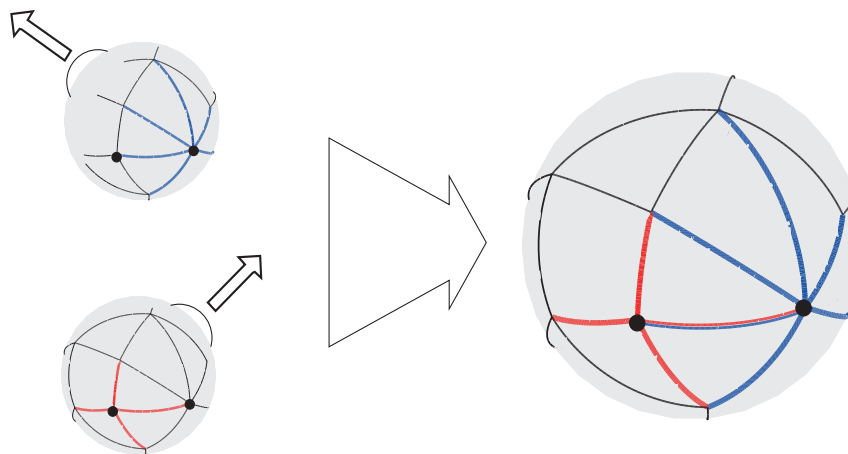
Miles Hansard

Graphs for 3D perception

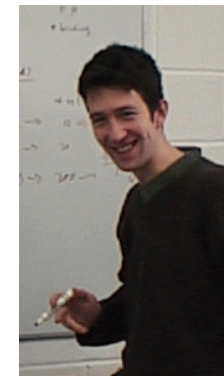
$$\Pi(s)$$



e.g. Tarr and Bülhoff (1998)
Glennerster, Hansard and Fitzgibbon (2001,2009)
Glennerster (2016)

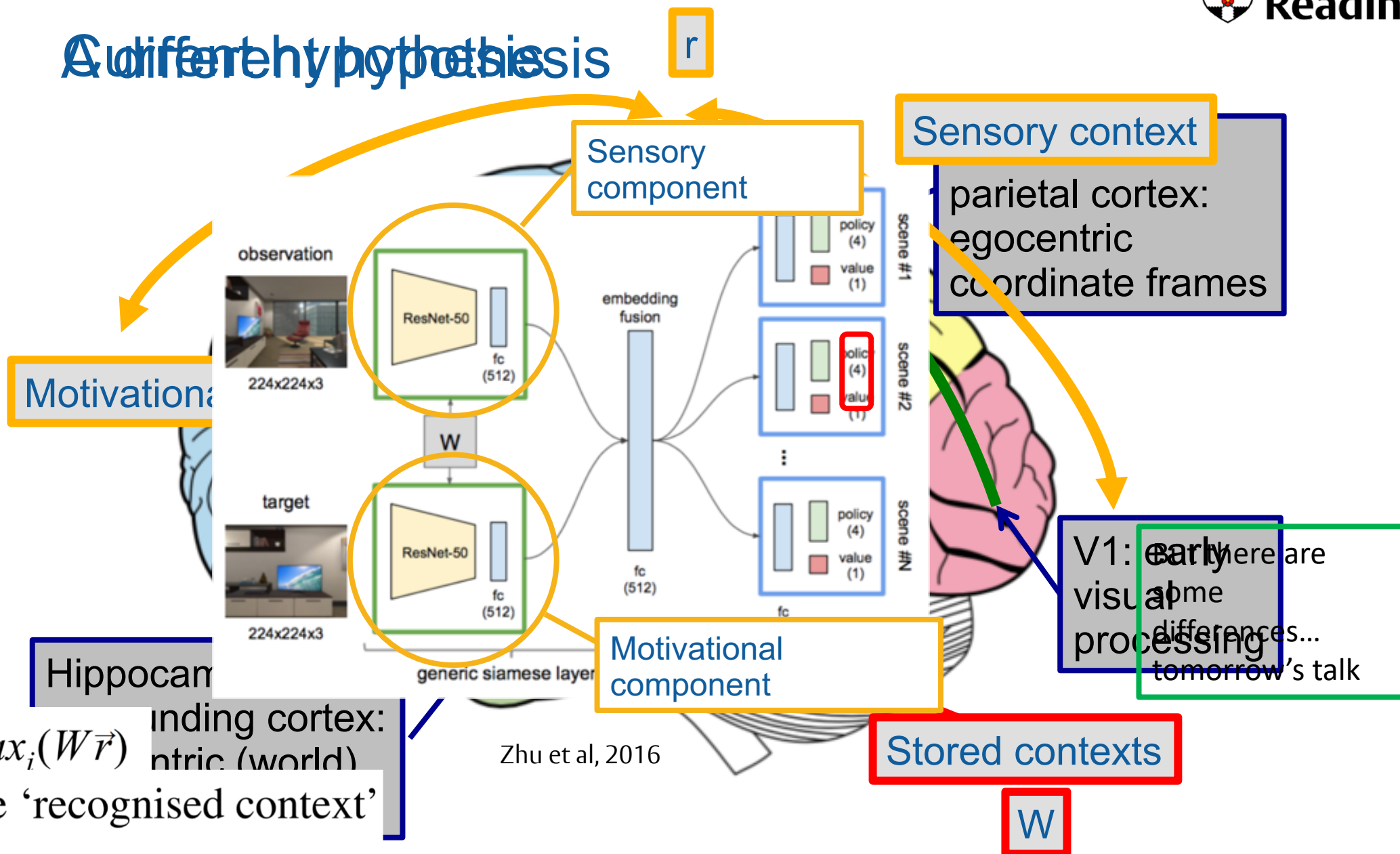


Information about object direction, distance, surface slant,
object shape with no 3D coordinates
Everything you need for a 2½-D sketch.



Andrew Fitzgibbon

A different hypothesis



Uniting different levels of spatial representation

Requires a longer conversation, but these elements open up the possibility of:

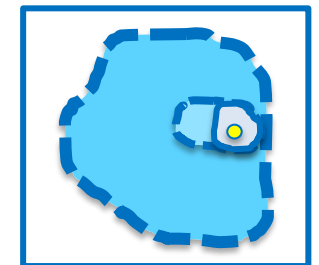
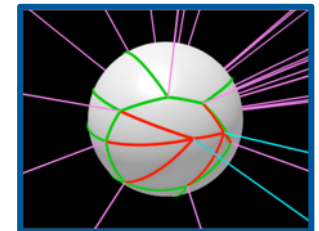
- A unified approach across many scales
 - fine scale detail, threading a needle
 - pointing to an unseen object
 - long range navigation
 - all can be related to a manifold of images rather than 3D coordinate frames
 - or, more ambitiously, a manifold of sensory+motivational states (so, including goals)
- Task-based
 - a ‘base camp’ representation is required to guide eye and head movements
 - but then judgements/actions can be computed on the fly. This explains apparent contradictions in human spatial representations for different tasks.

Collaborations

- UCLA
 - review on ‘human-like’ hierarchical tasks and spatial representation
- MIT
 - discussions about critical psychophysical experiments that could distinguish between predictions of physics engines and non-3D representations
- CMU
 - plan to make many cups of tea in VR (using AI2 THOR scenes in Unity) to compare generalization of behaviours in humans and RL
- Oxford
 - planned experiments to distinguish between the predictions of RL and other non-3D representations, e.g in interpolation between learned locations

Outline

- Evidence against 3D reconstruction
 - some briefly and
 - two examples in more detail
- What does the brain do instead?
 - a 2½-D sketch as ‘base camp’ for different tasks
 - could be implemented as a policy network
- Tomorrow
 - more on hierarchies of tasks
 - a different set of basis vectors for feature learning

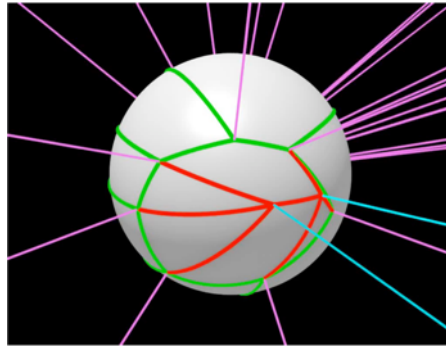




Jenny Vuong



Alex Muryy



Thanks...



Luise Gootjes-Dreesbach



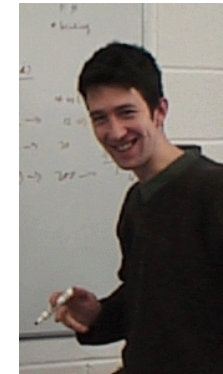
Peter Scarfe



James Stazicker



Miles Hansard



Andrew Fitzgibbon