Computational Cognitive Science

Lecture 16: High-level perception

Benjamin Peters

School of Informatics

University of Edinburgh

November 6, 2025

THE UNIVERSITY *of* EDINBURGH
informatics

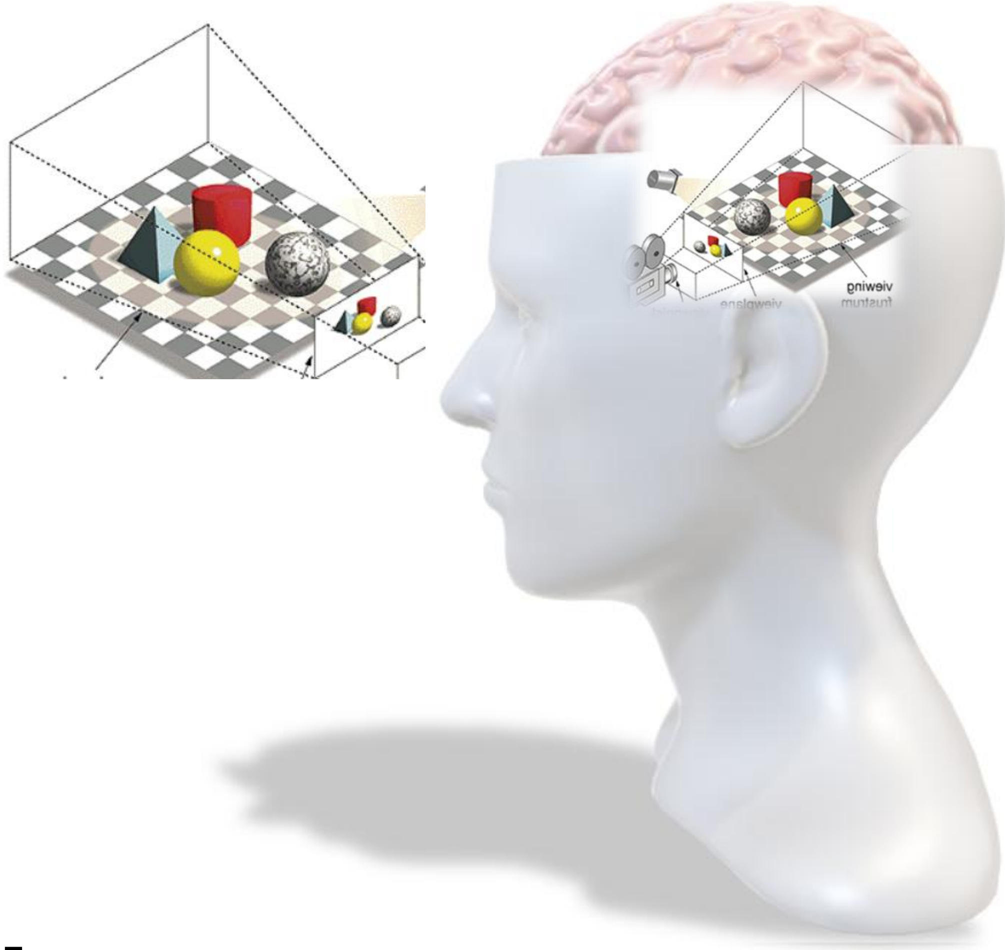
LVX VENIT IN MVNDVN ET DILEXERVNT HOMINES MAGIS TENEBRAS QVAM LVCEM. IO. 3. 19
ANTRVM PLATONICVM.
Maxima pars hominum cecis immersa tenebris
Volvitur assidue, et studio letatur inani:
Adspice ut obiectis obtutuis inhereat umbris,
Vt VERI simulacra omnes mirentur amentq,
Et stolidi vana ludantur imagine rerum.
Quam pauci meliore luto, qui in lumine puro
Secreti â stolidâ turbâ, ludibria cernunt
Rerum umbras rectaq; expendunt omnia lance:
Hi positâ erroris nebulâ dignoscere possunt
Vera bona, atque alios cecâ sub nocte latentes
Extrahere in claram lucem conantur, at illis
Nullus amor lucis, tanta est rationis egestas.
C.C. Harlemensis Inv.
Sanredam Sculpsit
Henr. Hondius excudit.
1604.
H.L. SPIEGEL FIGVRARI ET SCVLPI CVRAVIT. AC DOCTISS.o ORNATISS.oQZ D. PET. PAAW IN LVGDVN. ACAD. PROFESSORI MEDICO D.D.

Vision as unconscious inference

Ḥasan Ibn al-Haytham
(965-1040)

Hermann von Helmholtz
(1821-1894)

Analysis by synthesis



Bayesian brain

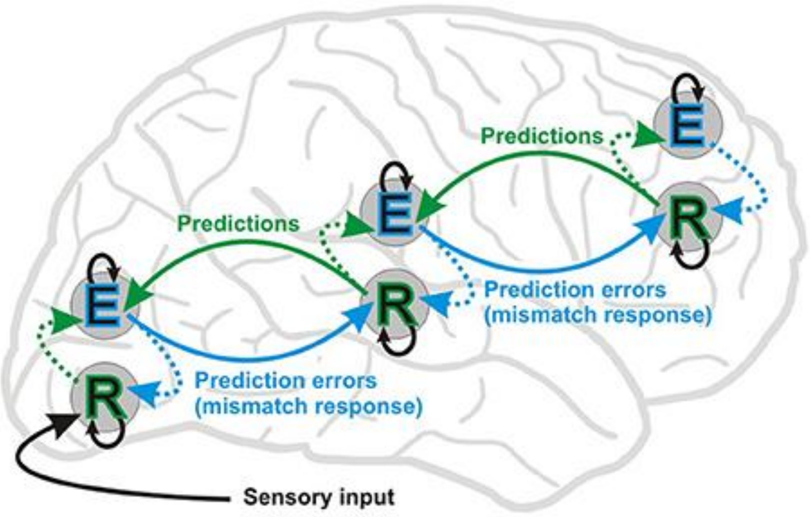
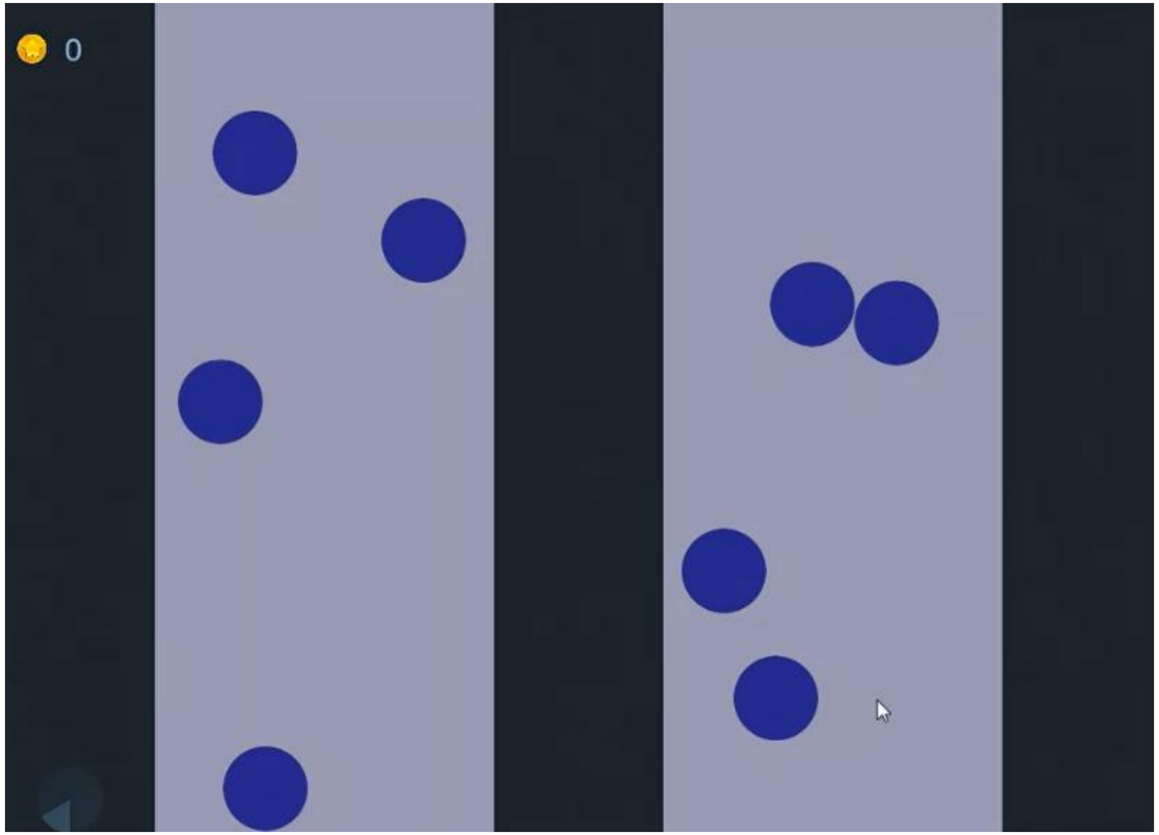


figure from Stefanics et al., 2014

Reading

- Vul, E., Alvarez, G., Tenenbaum, J. B., & Black, M. J. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. [link]

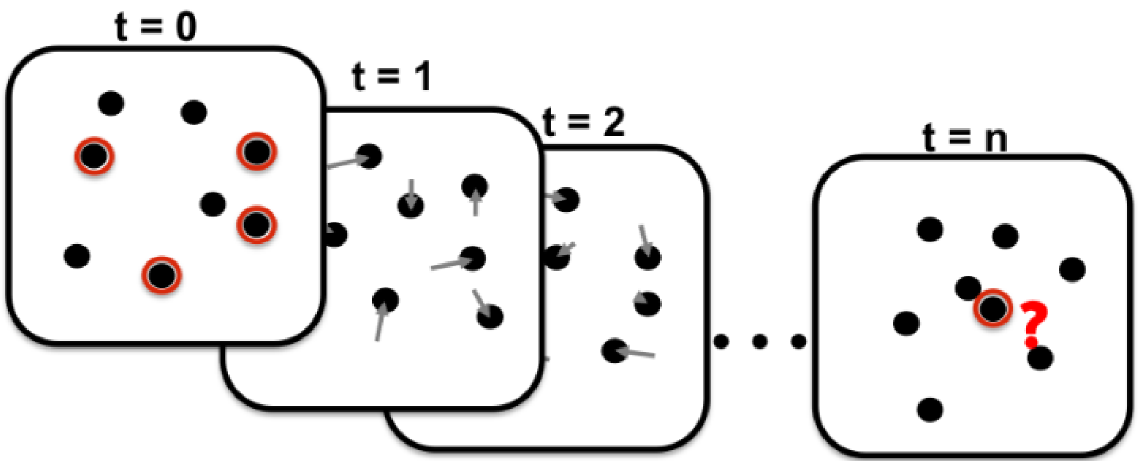

0

Cognitive capacity limitations during MOT

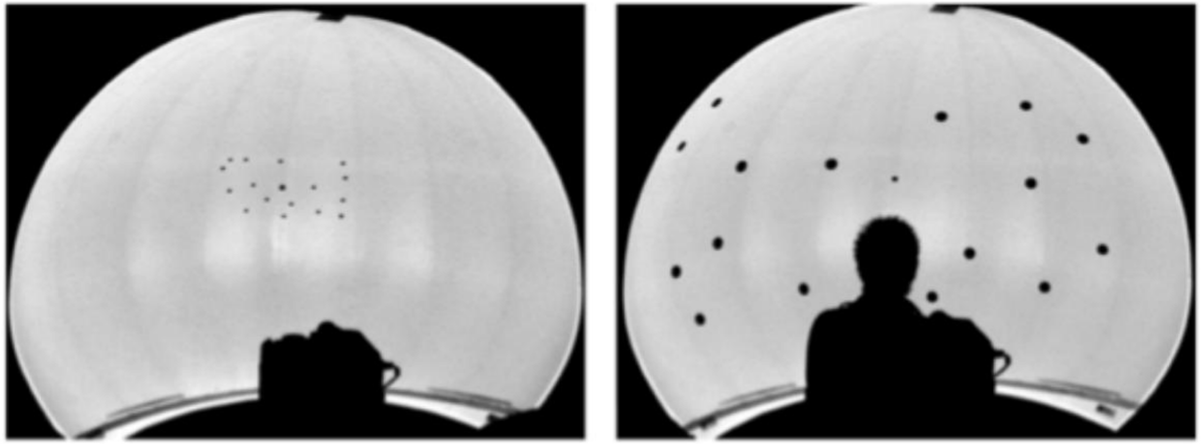
The ability to track multiple objects is highly capacity-limited: As more objects need to be tracked, tracking accuracy decreases. What limits our ability to track objects?

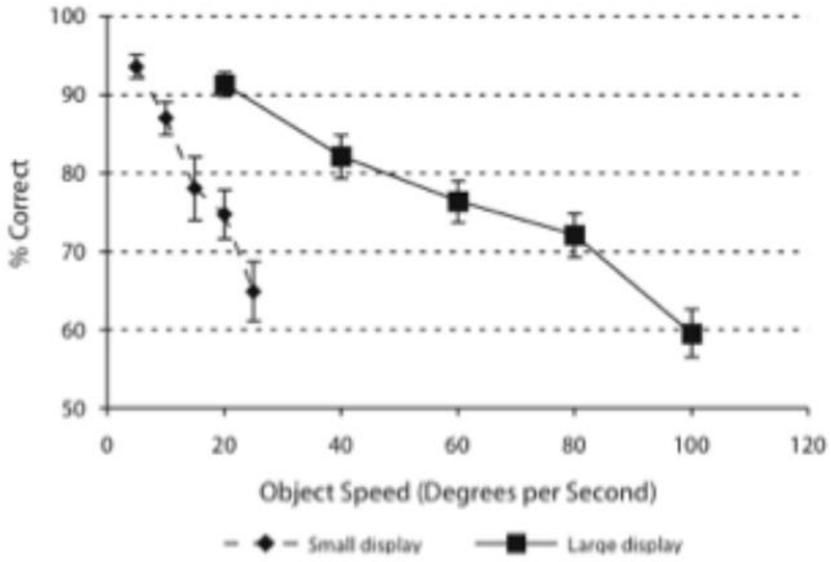
Several hypotheses have been entertained:

- Number of objects is limited (fixed slots hypothesis).
- Limitation on a flexible cognitive resource like attention or memory.
- What appears as a cognitive resource limitation, may be explained from the structure of the task itself and sensory noise.



- Accuracy decreases as more objects are tracked
- Accuracy increases if objects move slower
- Accuracy increases if objects are further apart from each other





Figures 1&2 from Vul et al., 2009

Model assumptions

- Human behaviour in a MOT task can be described by a Bayesian tracking framework
- Humans have a generative model of the objects motion and how they produce the observations.
- Optimal tracking: We assume that the humans' internal generative model is the same as the one that produced the observations.
- The evidence (observations) is then used to infer a posterior over the latent variables in the generative model (object position and speed)

Generative model for one object

$$x_t = x_{t-1} + v_t$$

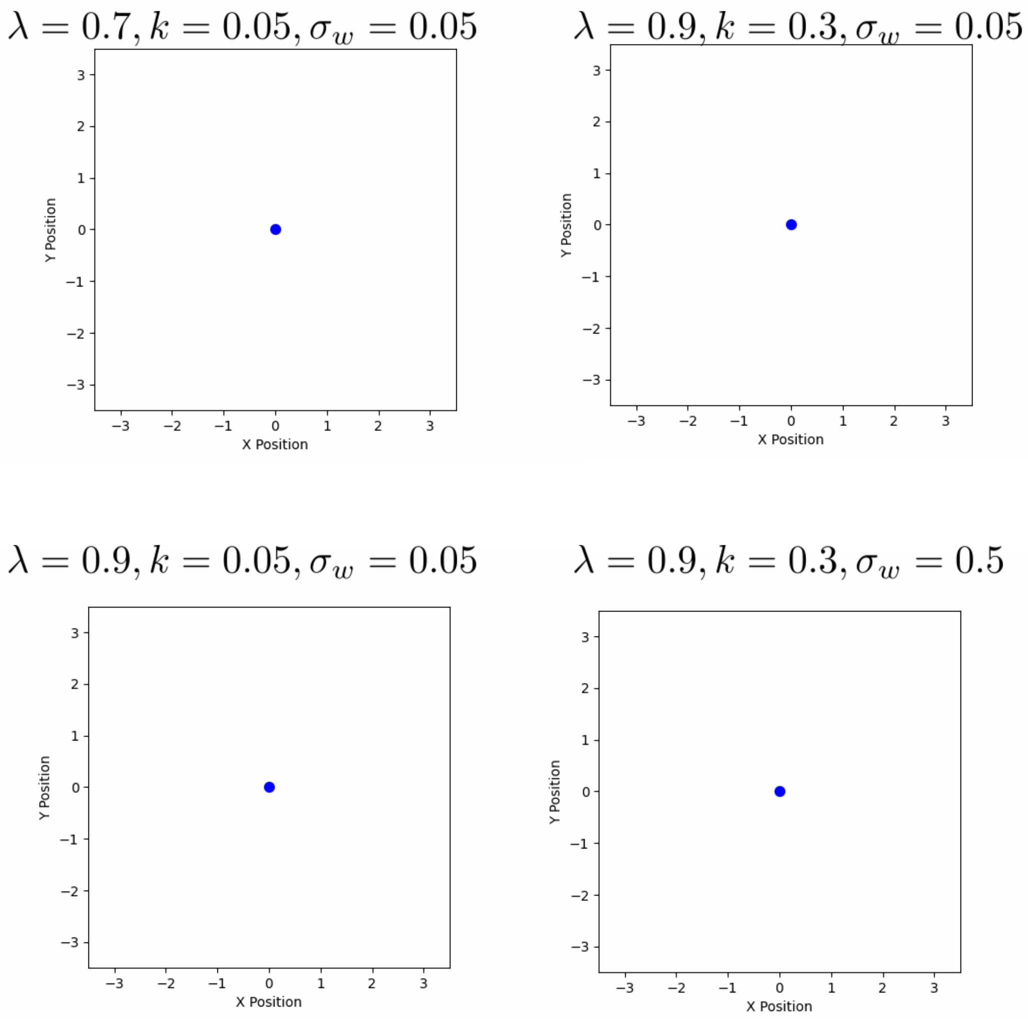
$$v_t = \lambda v_{t-1} - k x_{t-1} + w_t$$

$$w_t \sim N(0, \sigma_w)$$

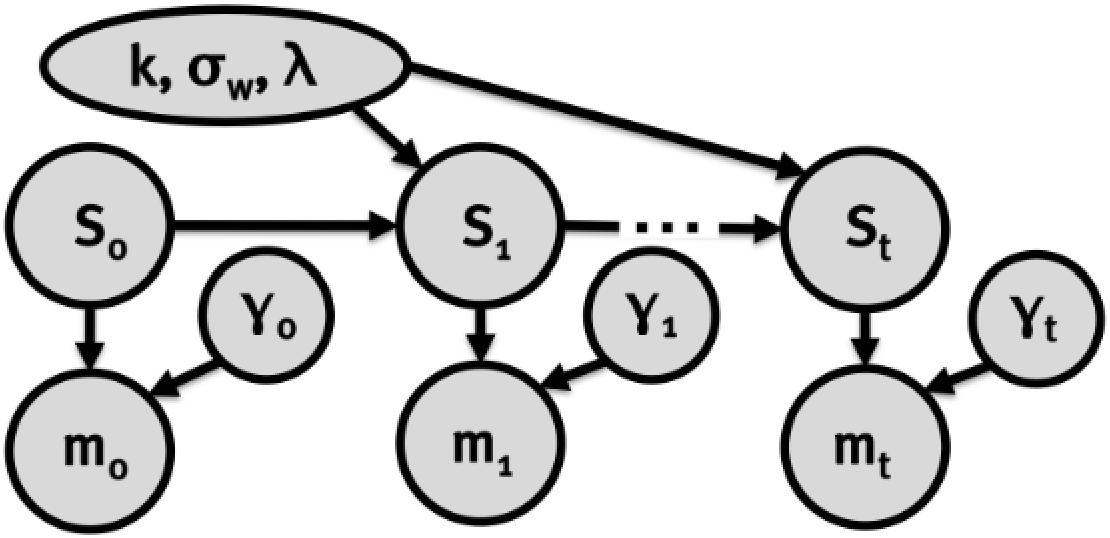
k : spring constant

λ : friction

Ornstein-Uhlenbeck process
("Brownian motion on a spring")



Generative Model



S_t: state of tracked objects (position and velocity)

γ_t: permutation vector

m_t: “unlabeled” bag of observations

Example

S_t

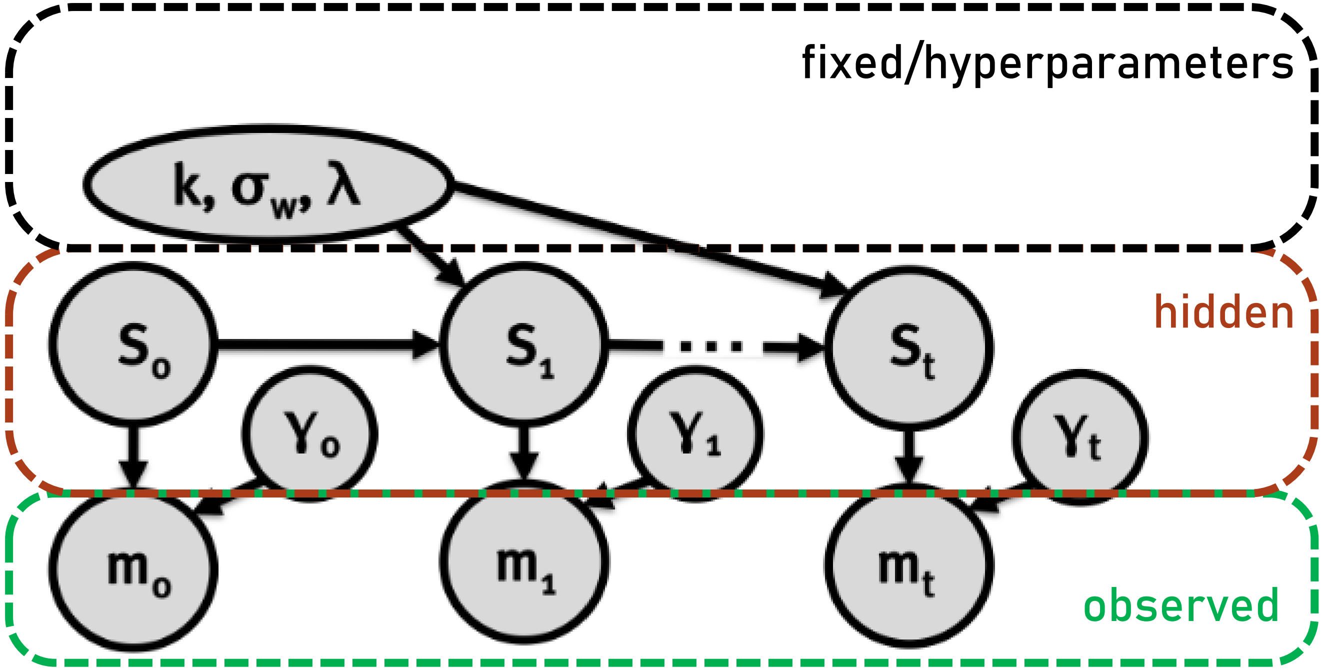
object1:
- position: [-0.3, 0.8]
- velocity: [0.05, -0.01]

object 2:
- position: [0.3, 0.1]
- velocity: [-0.04, 0.03]

γ_t [2, 1]

m_t [0.3, 0.1], [-0.04, 0.03]
[-0.3, 0.8], [0.05, -0.01]

Inference with the generative model



Given my all observations $\mathbf{m_0}$ until $\mathbf{m_t}$, what should my belief $\mathbf{S_t}$ be?

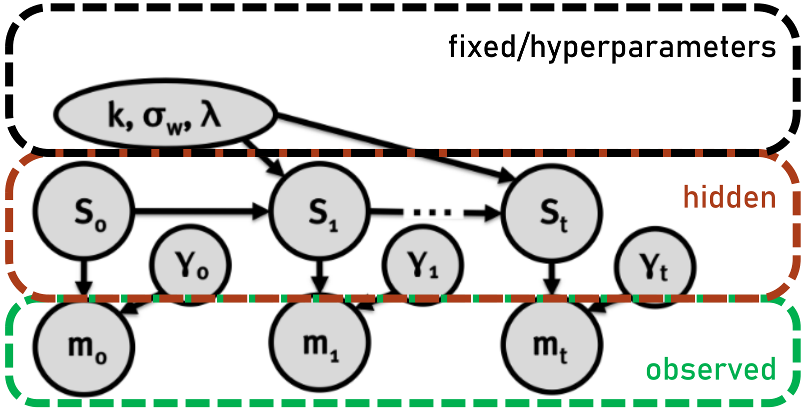
$$p(S_t, \gamma_t | m_{<t})$$

$\mathbf{S_t}$: state of tracked objects (position and velocity)

γ_t: permutation vector

$\mathbf{m_t}$: “unlabeled” bag of observations

Inference with the generative model



Given my all observations $\mathbf{m_0}$ until $\mathbf{m_t}$, what should my belief $\mathbf{S_t}$ be?

Markov condition $p(S_t, \gamma_t | m_{<t}) = p(S_t, \gamma_t | m_t, S_{t-1})$

-> Given my last belief about the position and velocity of the objects ($\mathbf{S_{t-1}}$) and the new observations $\mathbf{m_t}$, what should my new belief $\mathbf{S_t}$ be?

$\mathbf{S_t}$: state of tracked objects

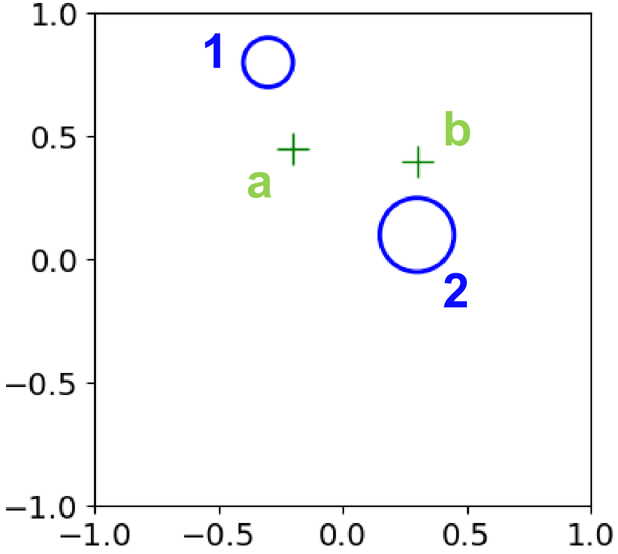
γ_t: permutation vector

$\mathbf{m_t}$: "unlabeled" bag of observations

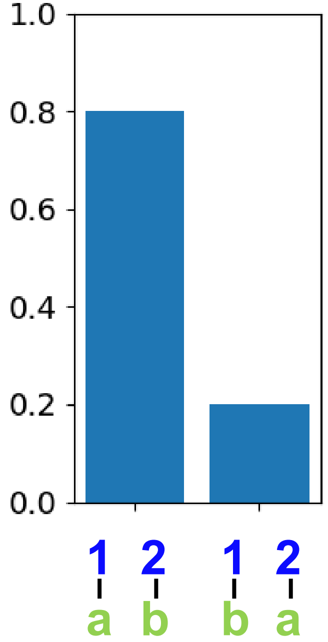
The posterior factorizes into two components:

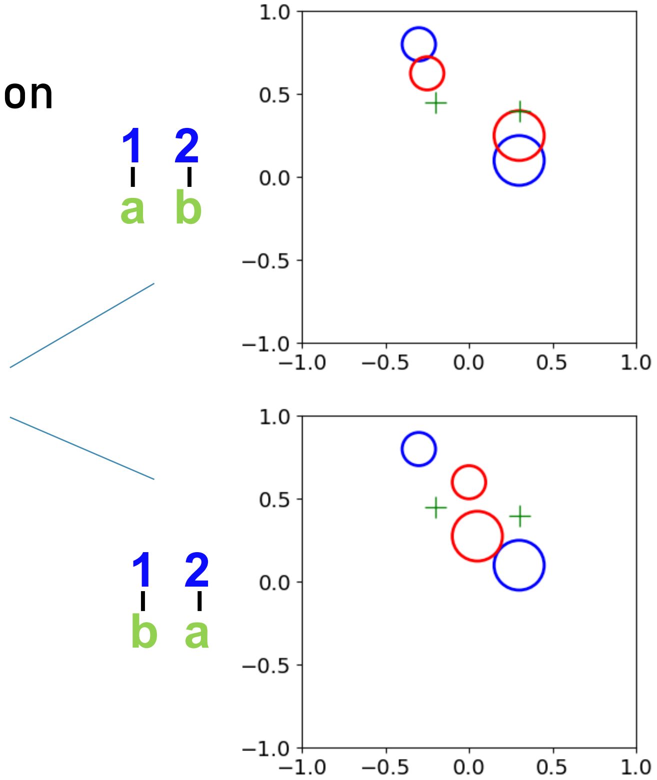
$$p(S_t, \gamma_t | m_t, S_{t-1}) = \underbrace{p(S_t | \gamma_t, m_t, S_{t-1})}_{\text{updated position/speed for each object given the assignments}} \underbrace{p(\gamma_t | m_t, S_{t-1})}_{\text{belief about assignment of observation to objects}}$$

Intuitions for the particle filter



proposal distribution





The posterior is approximated with a set of particles. Each particle is a hypothesis of $\mathbf{S_t}$ and $\boldsymbol{\gamma_t}$

Importance weight per particle: how representative is it for the posterior

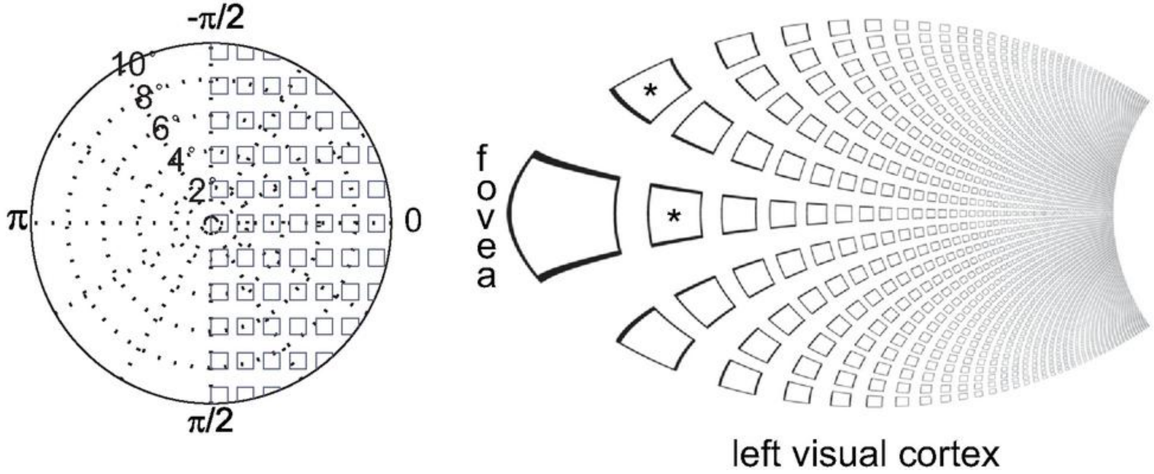
prediction: belief about object position/speed at time t based on the last belief S_{t-1}.

observation

updated beliefs S_t (via Kalman filter)

Observation likelihood $p(m_t|S_t, \gamma_t)$

- Perceptual uncertainty of <u>position</u>: linear increase in uncertainty with eccentricity, in line with cortical magnification.

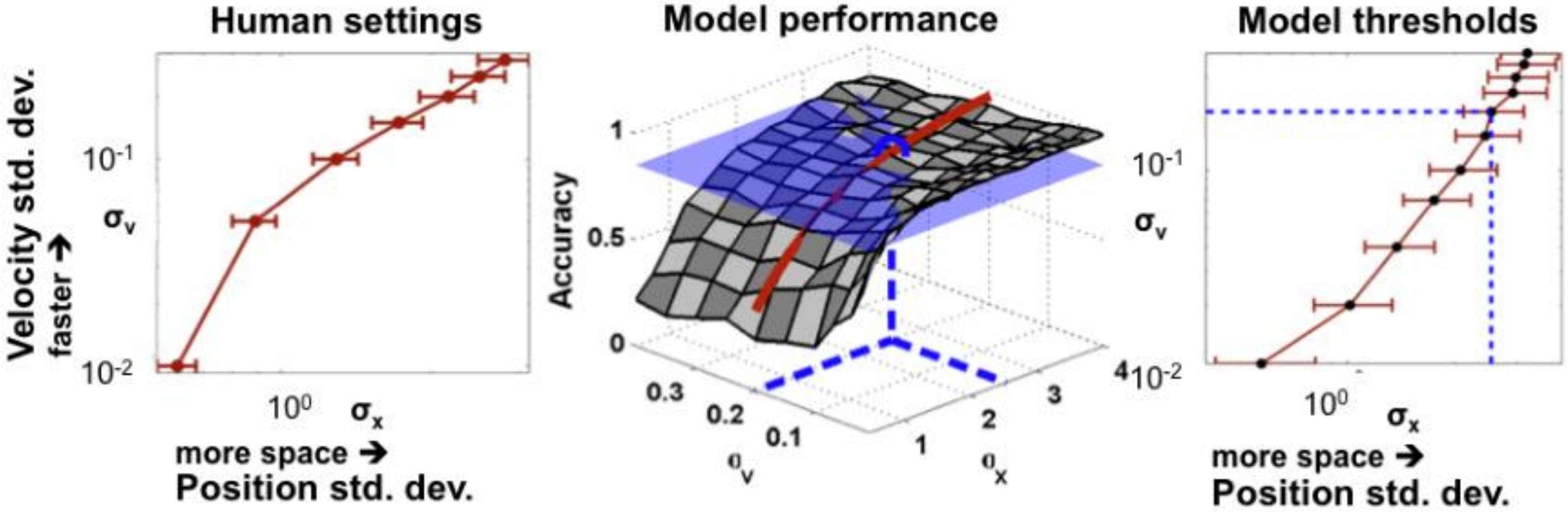


Schira et al. (2010)

- Perceptual uncertainty as a function of <u>speed</u>: as speed increases, uncertainty grows proportionally, in line with Weber's law.

Model predictions and results

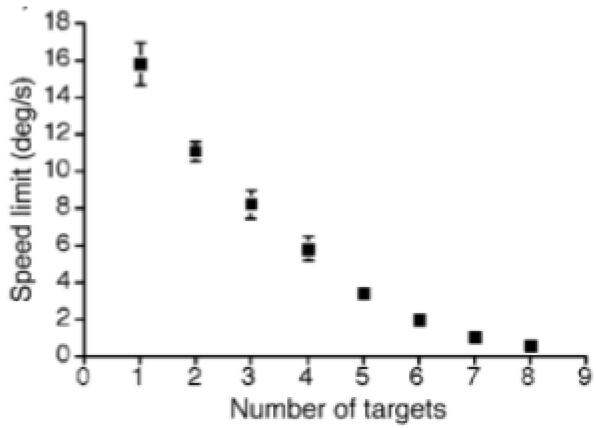
- Increasing speed and decreasing distance between objects both make the task harder.
- Faster speed: uncertainty in prediction increases
- Less spacing: confusions (wrong assignments) increase

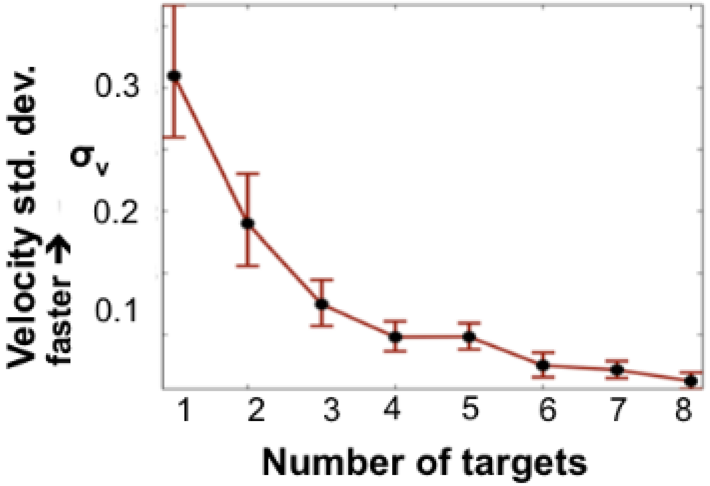


- Humans de/increasing distance between points to “comfortably” track 3 of 6 objects.

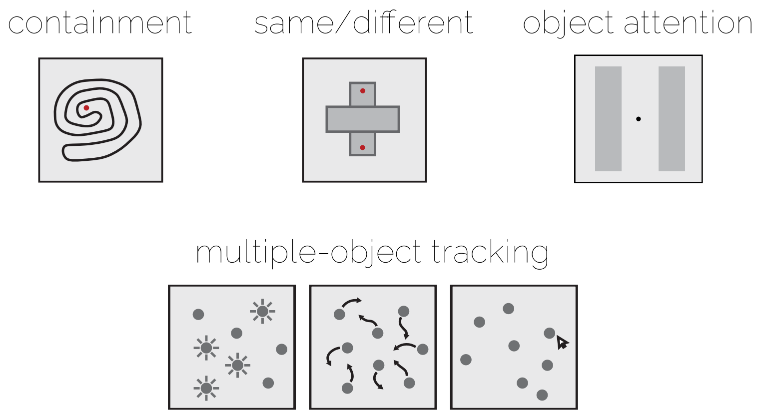
Cognitive limitations?

- Are spacing and speed effects on tracking accuracy the consequence of cognitive capacity limitations? Not necessarily, these effects are the consequence of the structure of the task and limits on the available perceptual information.
- But the model has to be augmented with a set-size dependent noise (attentional or memory limitations) to capture the tradeoff between number of targets and speed.



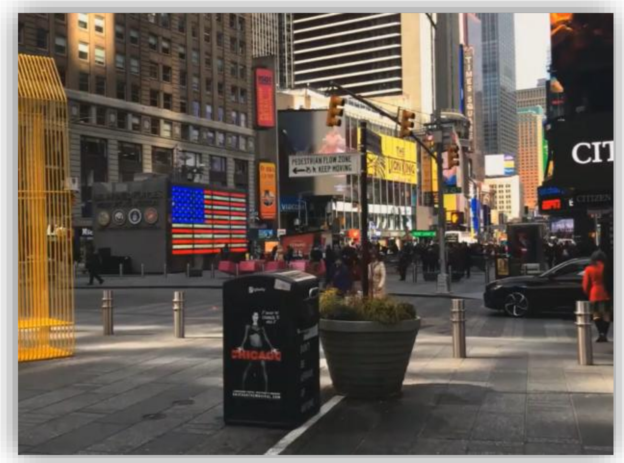


Traditional cognitive science

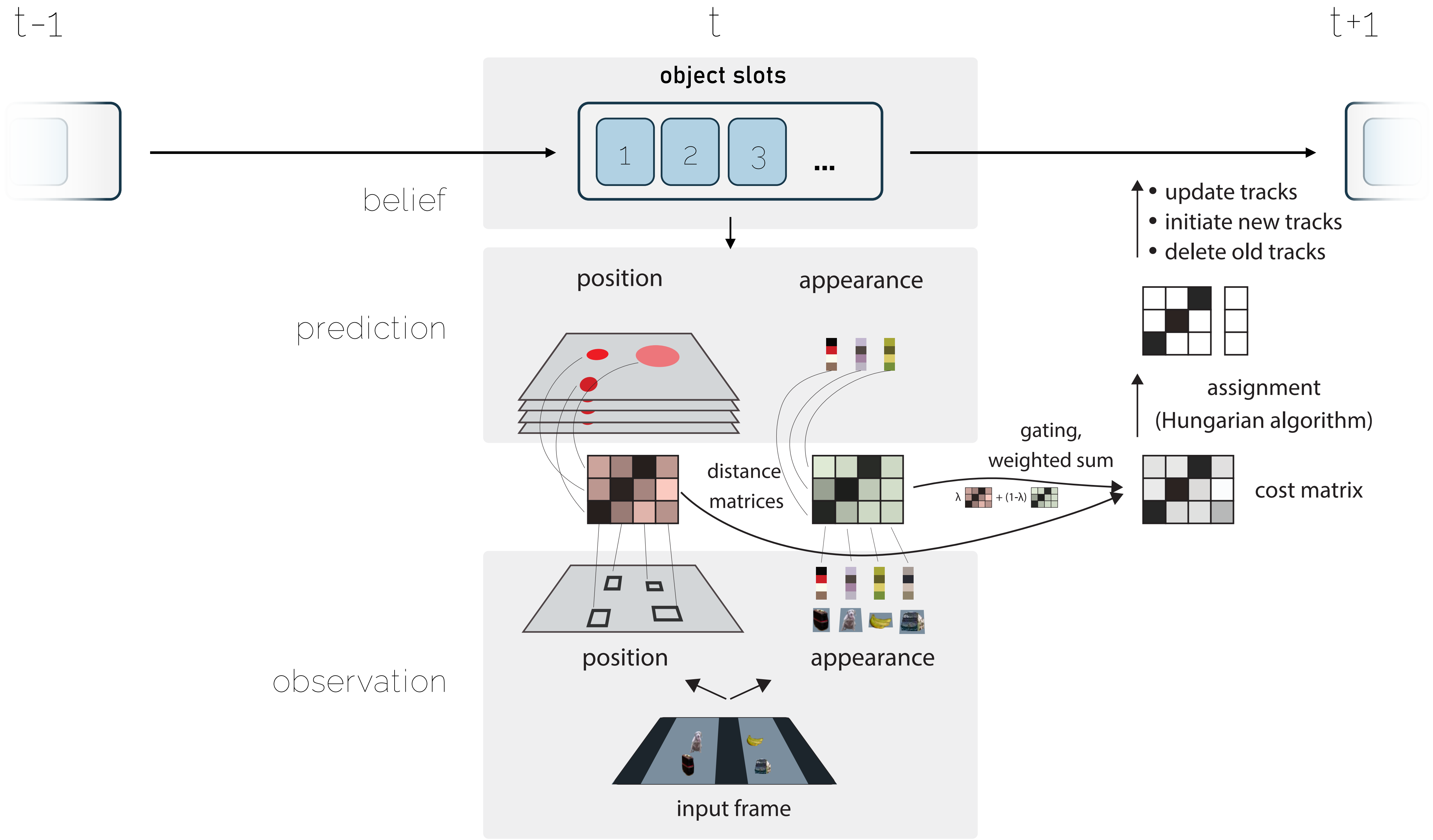


- highly abstracted toy tasks
- high experimental control

Real-world vision



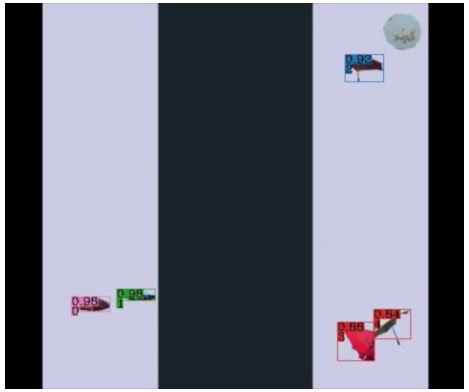
- real-world complexity, ecological validity
- less experimental control

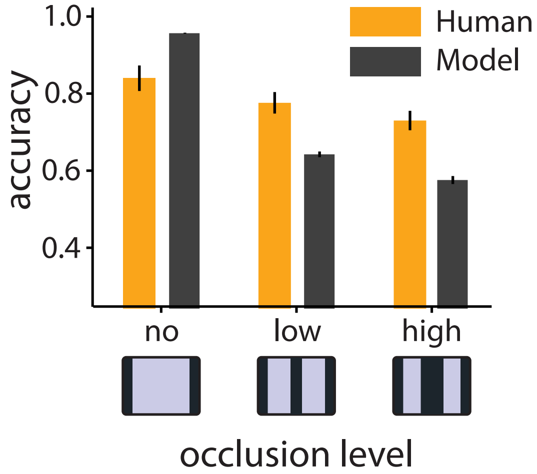


Peters, Butkus, & Kriegeskorte, 2022; Peters, Butkus, & Kriegeskorte, 2024

How do humans track multiple objects through occlusion?

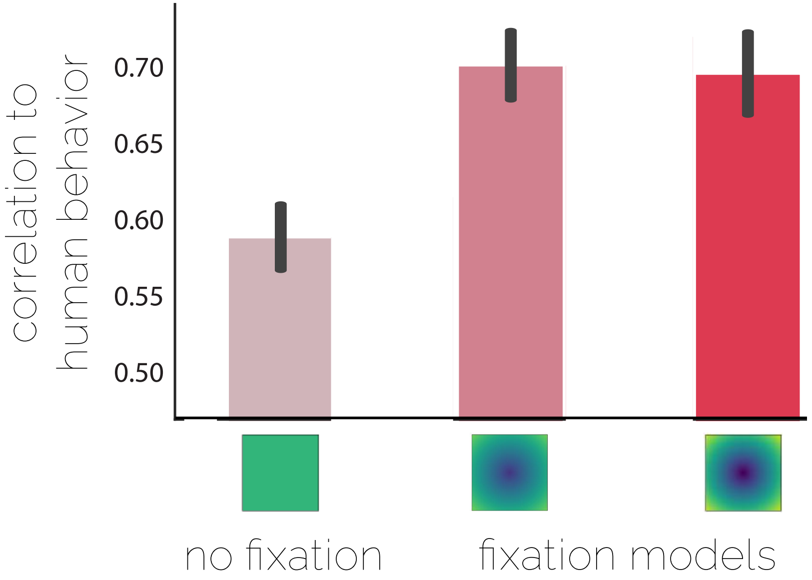
Humans vs. machines



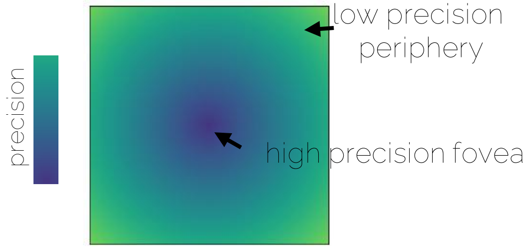


- models better than humans without occlusion
- models worse than humans for long occlusions

Gaze as determinant of performance



Modeling gaze improves the fit to human behaviour.



Peters, Butkus, & Kriegeskorte, 2022; Peters, Butkus, & Kriegeskorte, 2024

References

- Peters, B., Butkus, E., & Kriegeskorte, N. (2024). Human-like multiple object tracking through occlusion via gaze-following. *Proceedings of The 2nd Gaze Meets ML Workshop*, 181–196. https://proceedings.mlr.press/v226/peters24a.html
- Peters, B., Butkus, E., & Kriegeskorte, N. (2022). *How do humans and machine learning models track multiple objects through occlusion?* SVRHM 2022 Workshop @ NeurIPS. https://openreview.net/forum?id=n5OTU5qcj05
- Peters, B., & Kriegeskorte, N. (2021). Capturing the objects of vision with neural networks. *Nature Human Behaviour*, *5*(9), Article 9. https://doi.org/10.1038/s41562-021-01194-6
- Schira, M. M., Tyler, C. W., Spehar, B., & Breakspear, M. (2010). Modeling Magnification and Anisotropy in the Primate Foveal Confluence. *PLOS Computational Biology*, *6*(1), e1000651. https://doi.org/10.1371/journal.pcbi.1000651
- Stefanics, G., Kremláček, J., & Czigler, I. (2014). Visual mismatch negativity: A predictive coding view. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00666
- Vul, E., Alvarez, G., Tenenbaum, J. B., & Black, M. J. (2009). Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 1955–1963). Curran Associates, Inc. http://papers.nips.cc/paper/3828-explaining-human-multiple-object-tracking-as-resource-constrained-approximate-inference-in-a-dynamic-probabilistic-model.pdf