KTH ROYAL INSTITUTE
OF TECHNOLOGY

What is Embodiment, and How Does It Affect the Way We Function?

Hedvig Kjellström

Professor of Computer Science

Head of the Department of **Robotics, Perception, and Learning**

What is Embodiment?

Here, in the Cognitive Psychology sense (situatedness, to have a physical location and form in the world)

How Does It Affect The Way We Function?

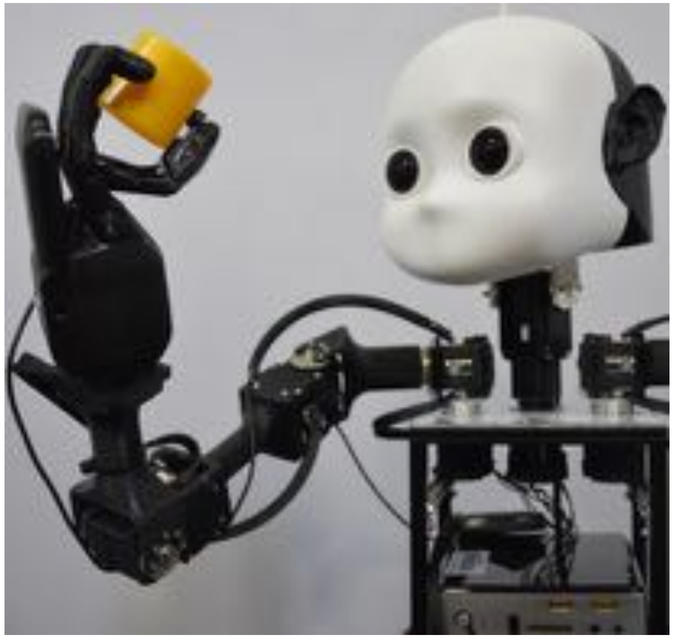
Embodied Cognition

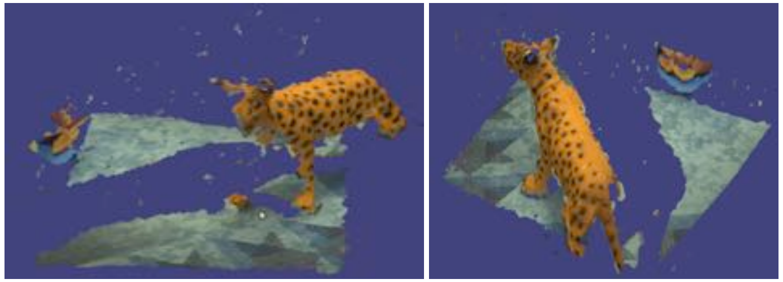
Three aspects from an AI perspective

M. V. Butz and E. F. Kutter. *How the Mind Comes into Being*, 2017

Aspect 1:
Perception-Action Loop

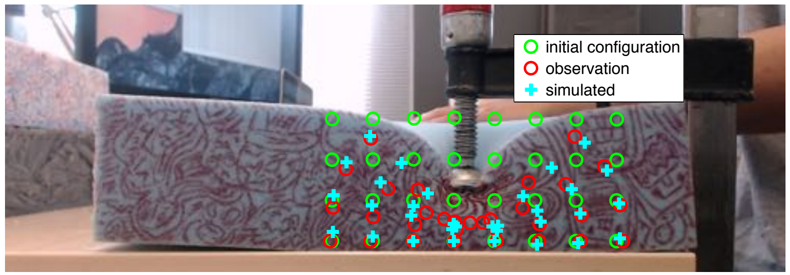
Active Perception – Getting More Info Through Embodiment





(Jeannette Bohg, 2011)

(Niklas Bergström, 2011)



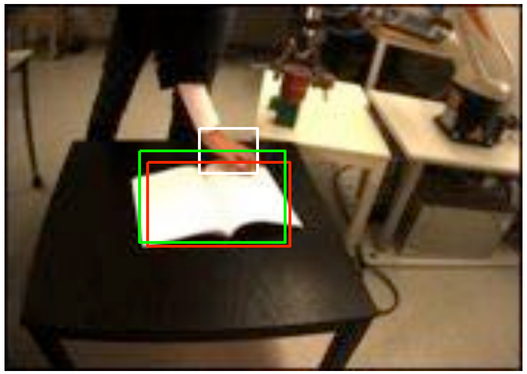
(Püren Güler, 2015)

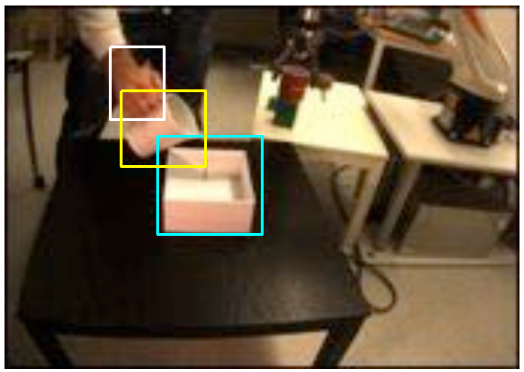
Embodiment Changes What We Need To Perceive – Object Affordances

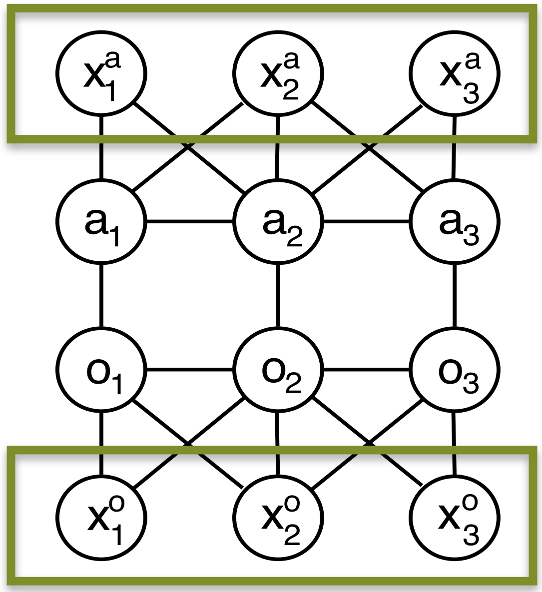
After (Bülthoff and Bülthoff, 2003)

M. V. Butz and E. F. Kutter. *How the Mind Comes into Being*, 2017
J. J. Gibson. *The Ecological Approach to Visual Perception*, 1979

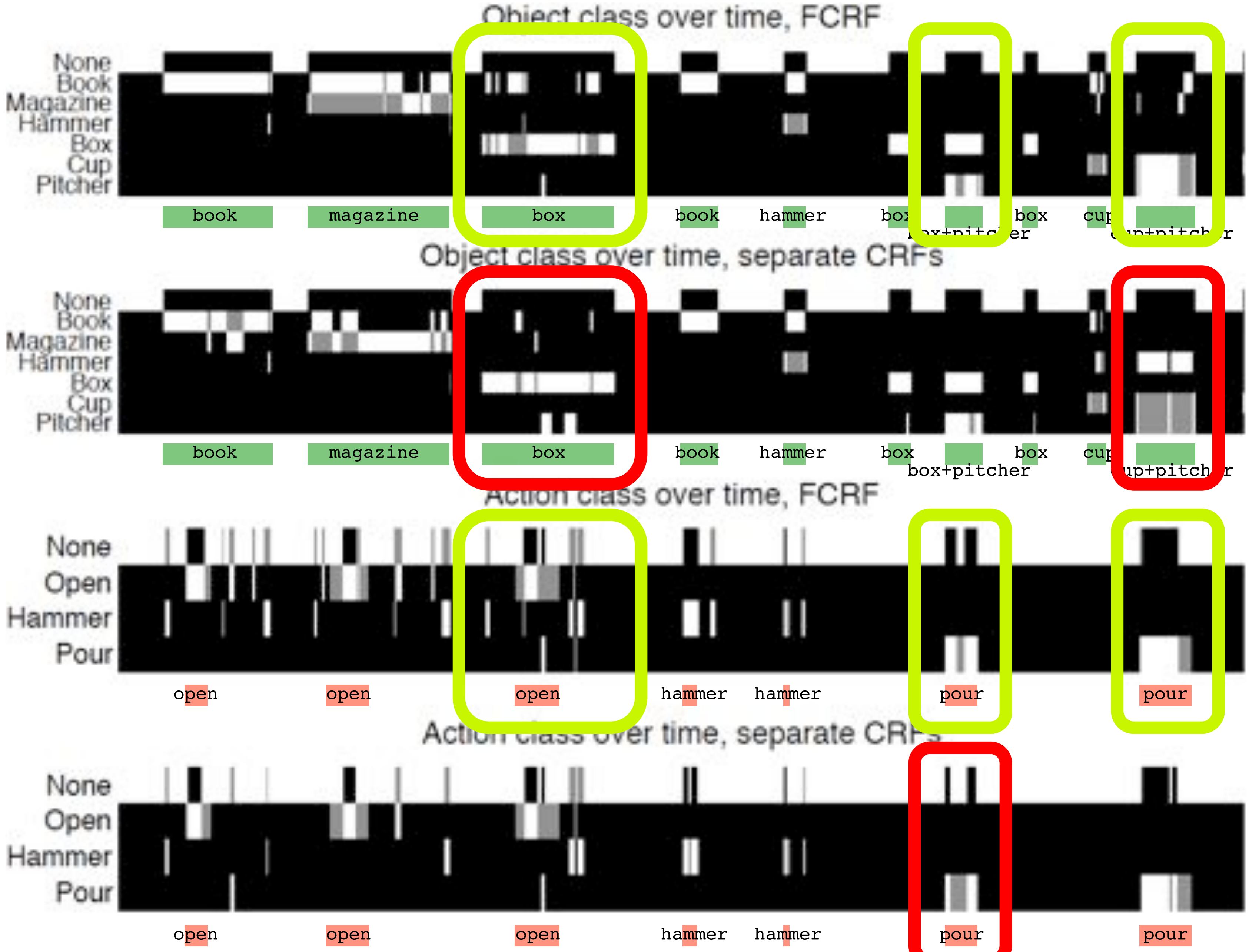
Object-Action (~Affordance) Recognition







H. Kjellström, J. Romero, and D. Kragic. Visual object-action recognition: Inferring object affordances from human demonstration. *Computer Vision and Image Understanding*, 115:81-90, 2011



H. Kjellström, J. Romero, and D. Kragic. Visual object-action recognition: Inferring object affordances from human demonstration. *Computer Vision and Image Understanding*, 115:81-90, 2011

Aspect 2:
Low Communication Bandwidth

Humans are Good at Communicating with Others – Artificial Systems Need to Be

Why is Human Communication Hard?

Embodiment factor

$$E = \frac{\text{computing power}}{\text{communication bandwidth}}$$

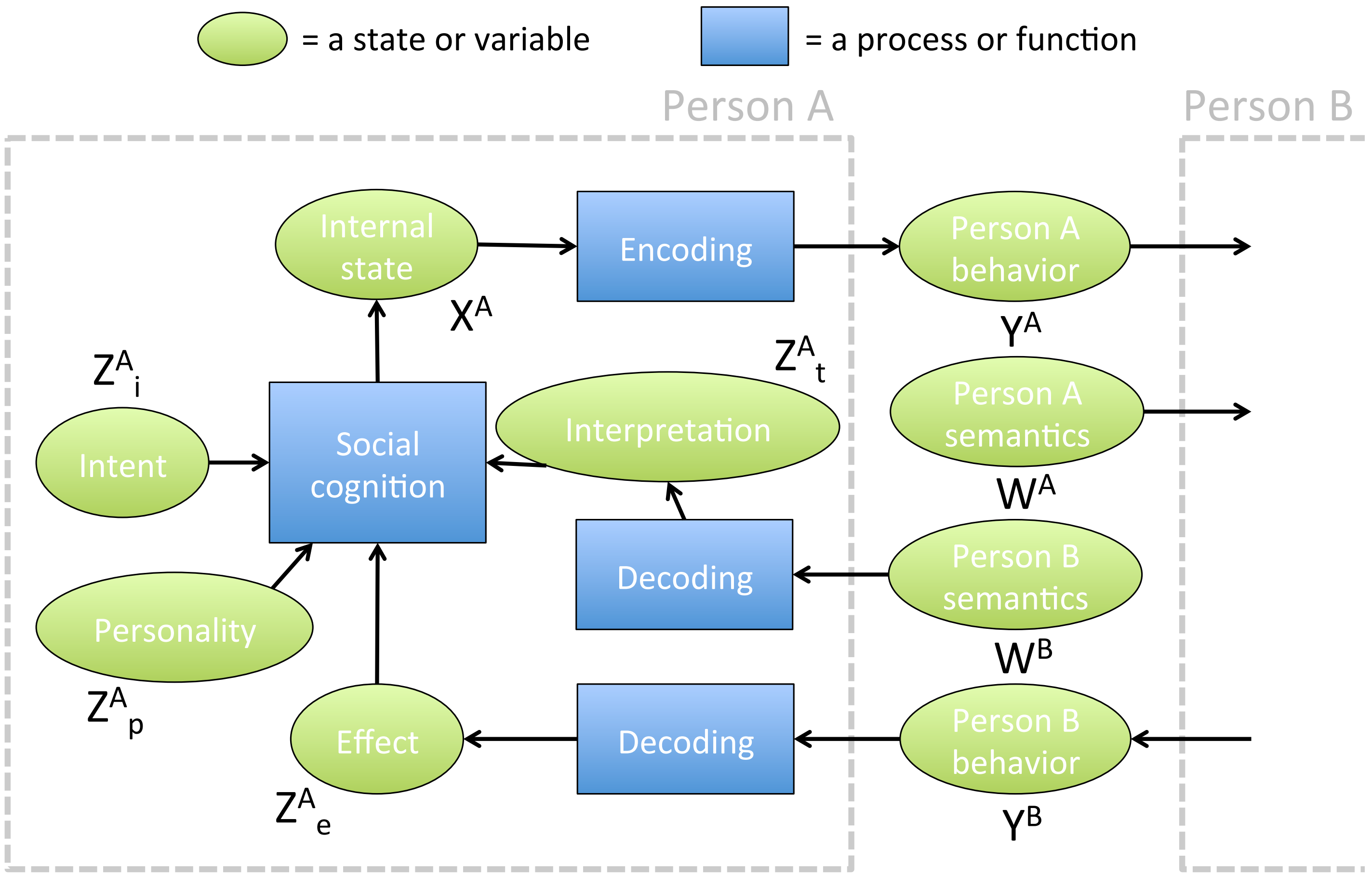
Human: $E \approx 10^{16}$

Computer: $E \approx 10$

Conclusions

1. Embodiment makes understanding hard
2. Need to *emulate* embodiment in artificial agent to enable understanding

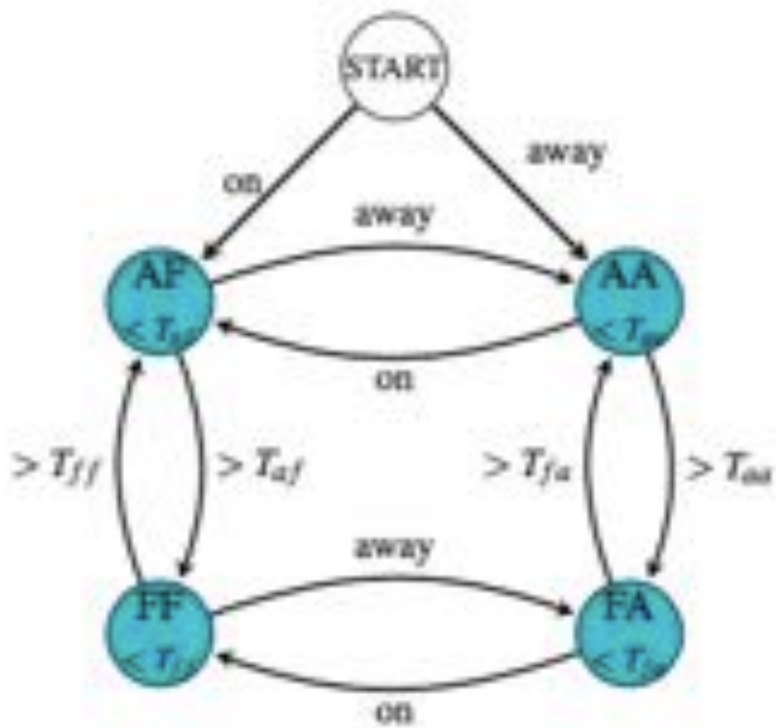
Generative Probabilistic Framework for Social Signal Processing



Adapted from D. MacKay. Formal analysis of communicative processes. In *Non-Verbal Communication*, pp 16-21, 1972

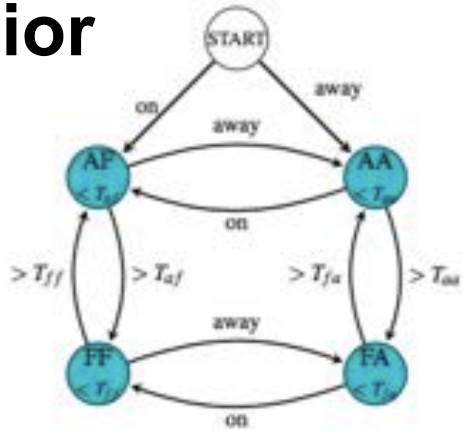
Perception and Production of Gaze Aversion Behavior

Yanxia Zhang
PostDoc 2016

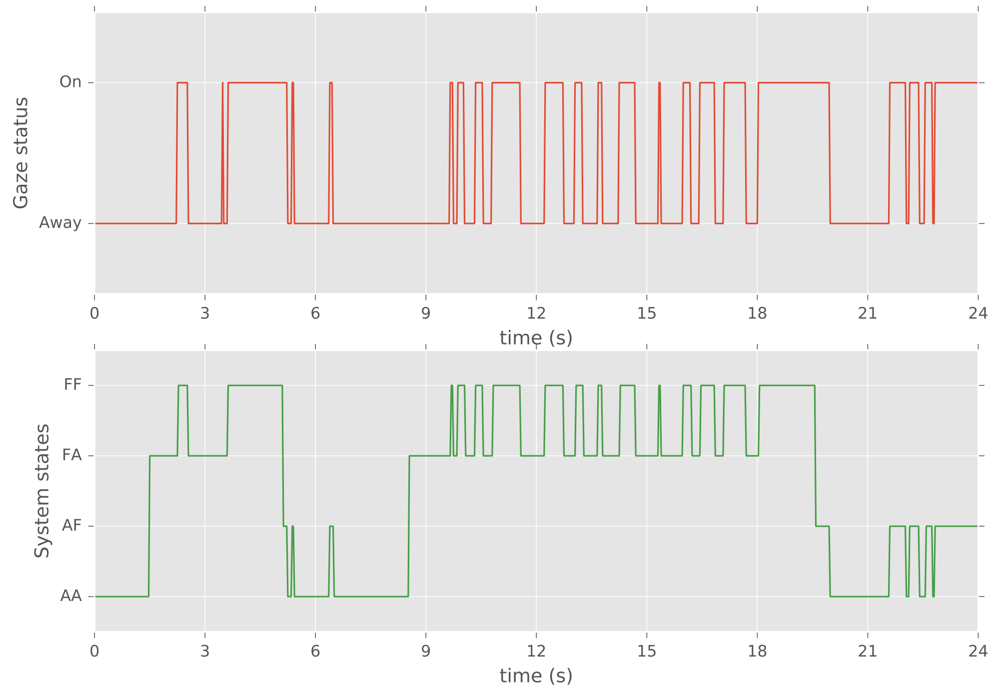


Y. Zhang, J. Beskow, and H. Kjellström. Look but don't stare: Mutual gaze interaction in social robots. *International Conference on Social Robotics*, 2017

Perception and Production of Gaze Aversion Behavior

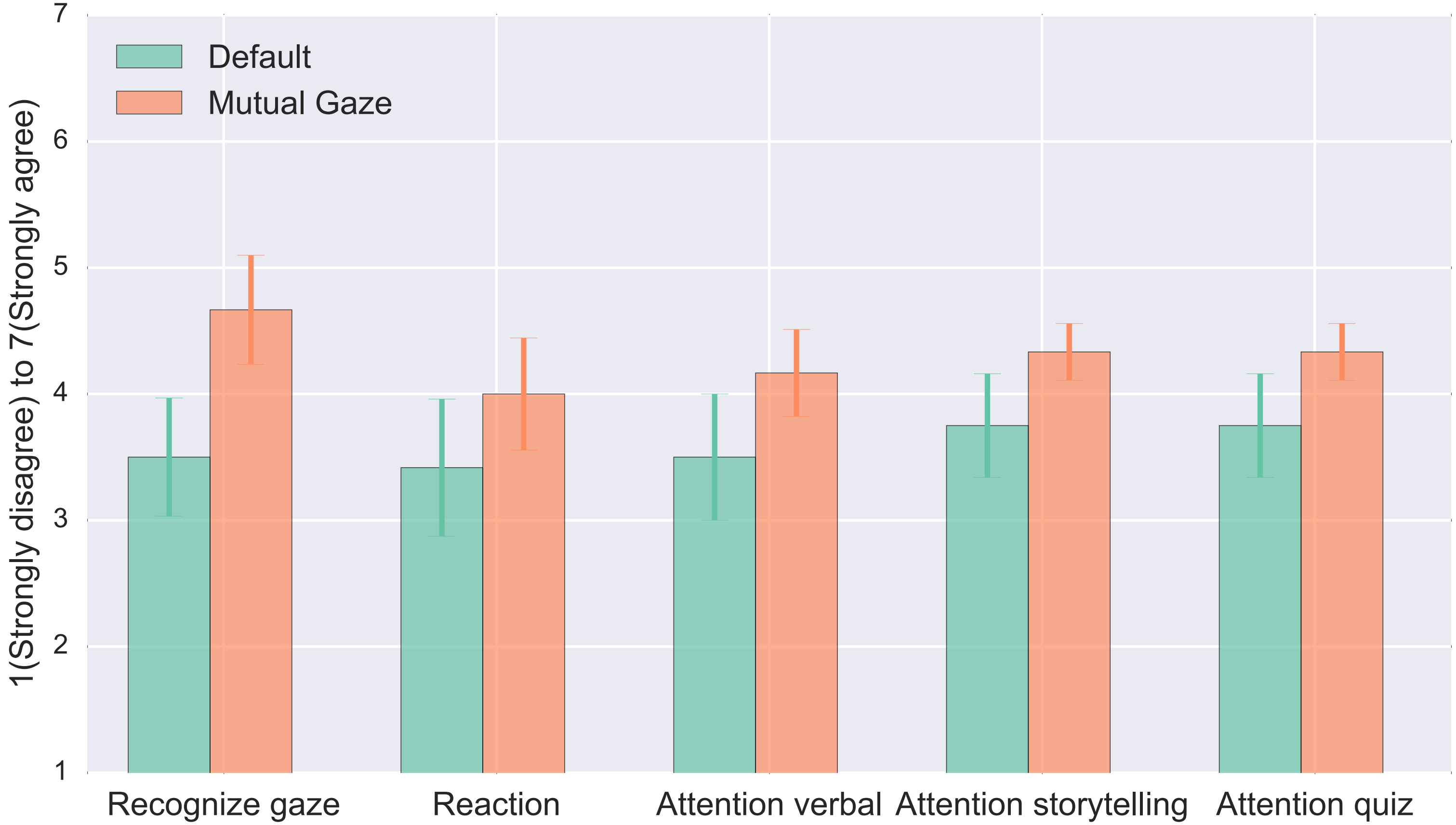


Yanxia Zhang
PostDoc 2016



Y. Zhang, J. Beskow, and H. Kjellström. Look but don't stare: Mutual gaze interaction in social robots. *International Conference on Social Robotics*, 2017

Perception and Production of Gaze Aversion Behavior

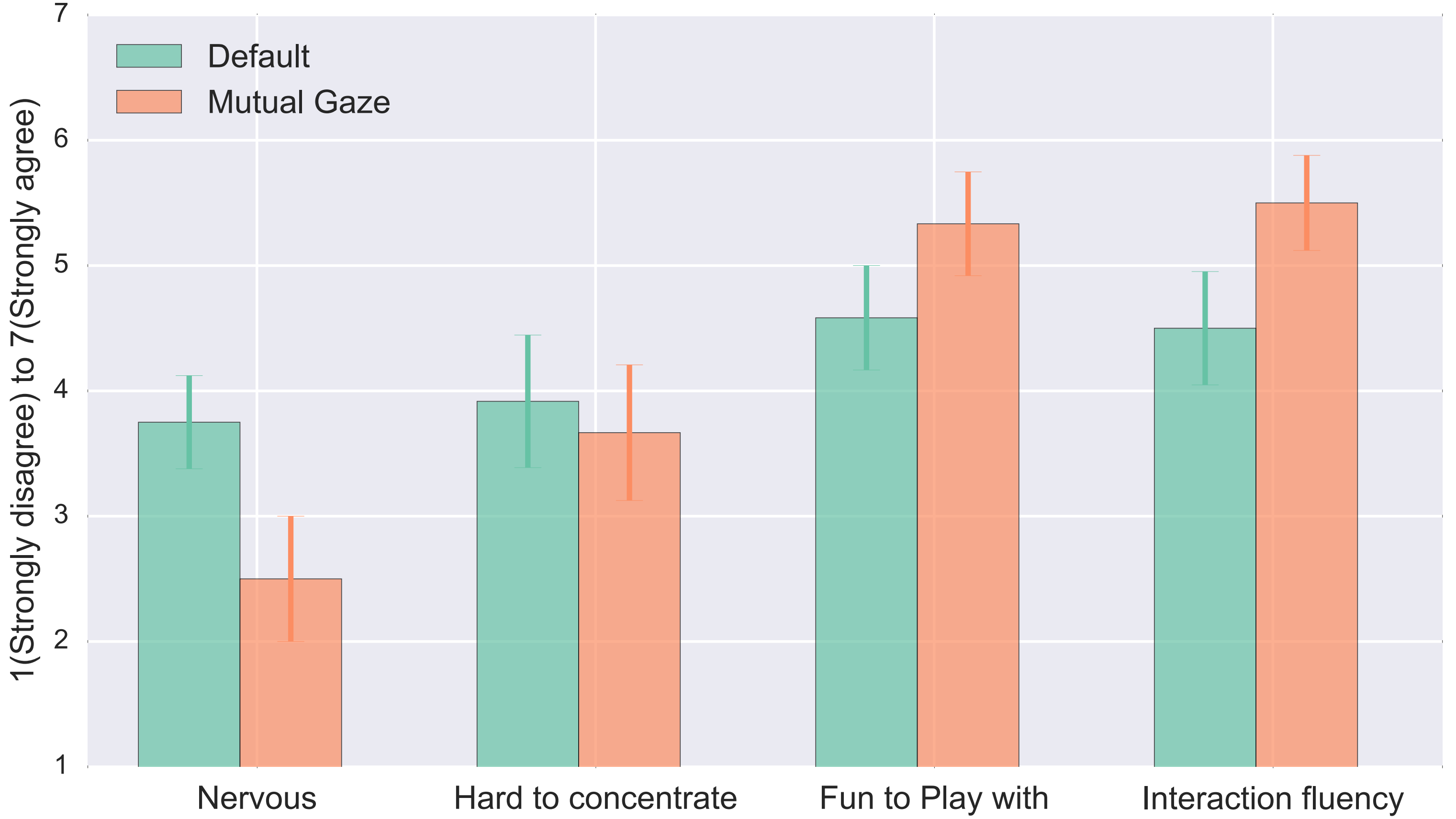


Yanxia Zhang
PostDoc 2016

Y. Zhang, J. Beskow, and H. Kjellström. Look but don't stare: Mutual gaze interaction in social robots. *International Conference on Social Robotics*, 2017

Perception and Production of Gaze Aversion Behavior

Yanxia Zhang
PostDoc 2016

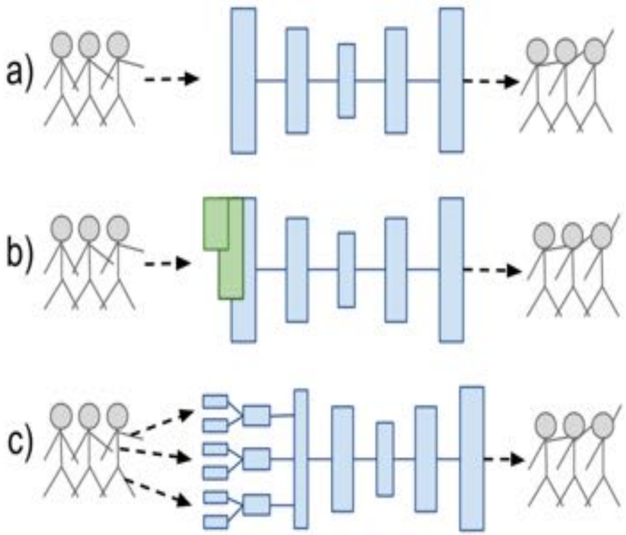


Y. Zhang, J. Beskow, and H. Kjellström. Look but don't stare: Mutual gaze interaction in social robots. *International Conference on Social Robotics*, 2017

Predicting Motion of Human Collaborator

Judith Bütepage
Joint with Danica Kragic

Fully-connected temporal encoder-decoders – generative models of human motion
Input: past motion (mocap)   Output: future motion



Symmetric temporal encoder

Convolutional temporal encoder

Hierarchical temporal encoder

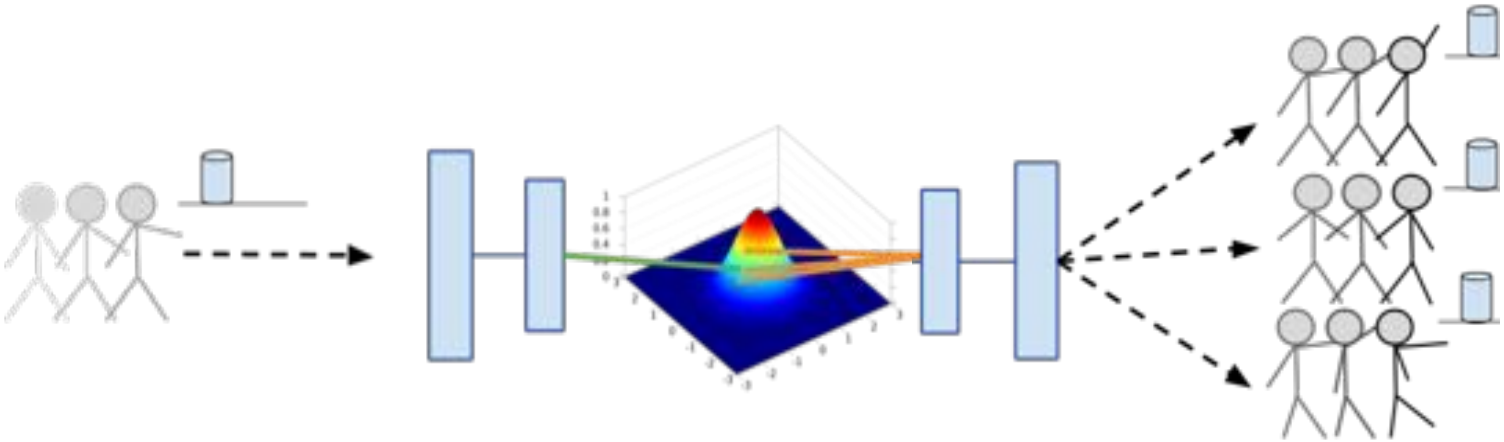
J. Bütepage, M. Black, D. Kragic, and H. Kjellström. Deep representation learning for human motion prediction and classification. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017

Predicting Motion of Human Collaborator

Judith Bütepage
Joint with Danica Kragic

Solution:

Conditional variational autoencoders [1] trained on skeletal data obtained from **RGB depth images**



[1] K. Sohn, H. Lee, and X. Yan, Learning structured output representation using deep conditional generative models. *NIPS*, 2015

J. Bütepage, H. Kjellström, and D. Kragic. Anticipating many futures: Online human motion prediction and synthesis for human-robot collaboration. *IEEE International Conference on Robotics and Automation*, 2018

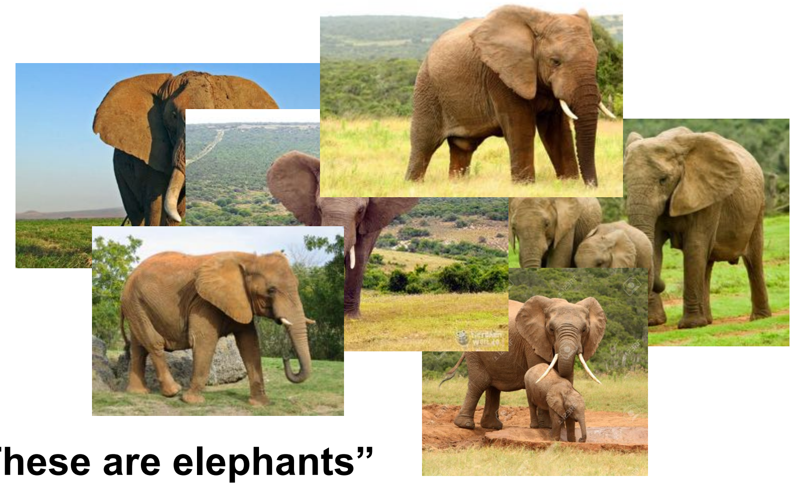
Aspect 3:
Learning from Few Examples

Humans are Good at Continuous and Dynamic Learning – Artificial Systems Need to Be

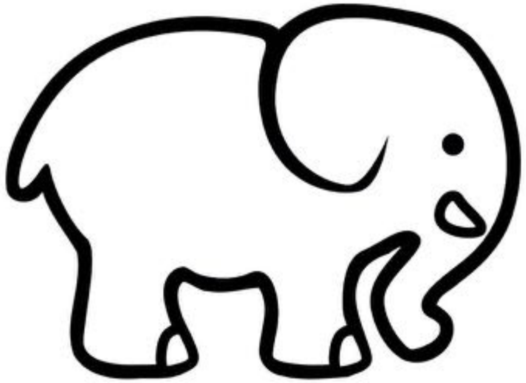
Embodiment Shapes the Way We Learn – Learning from Few Examples

State of the art ML algorithm

Toddler



”These are elephants”



”This is a drawing of an elephant”

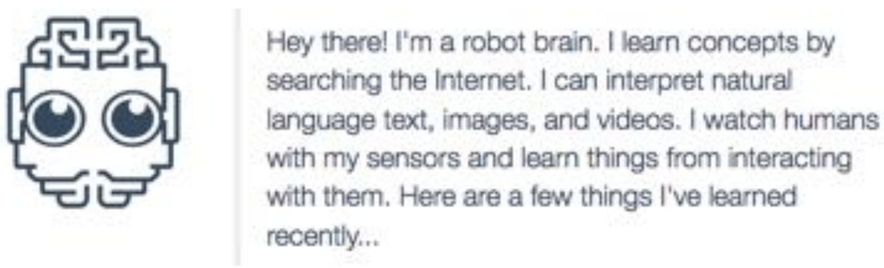
”This is an elephant!”

B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. Building machines that learn and think like people. *Behavioral and Brain Sciences* 24:1-101, 2016

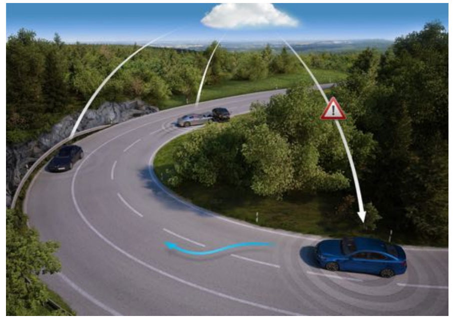
Embodiment Shapes the Way We Learn – But Still Learn from Many Examples?

Alternative strategy – provide enough training data!

Crowd Sourcing



The Robo Brain project (http://robobrain.me/)

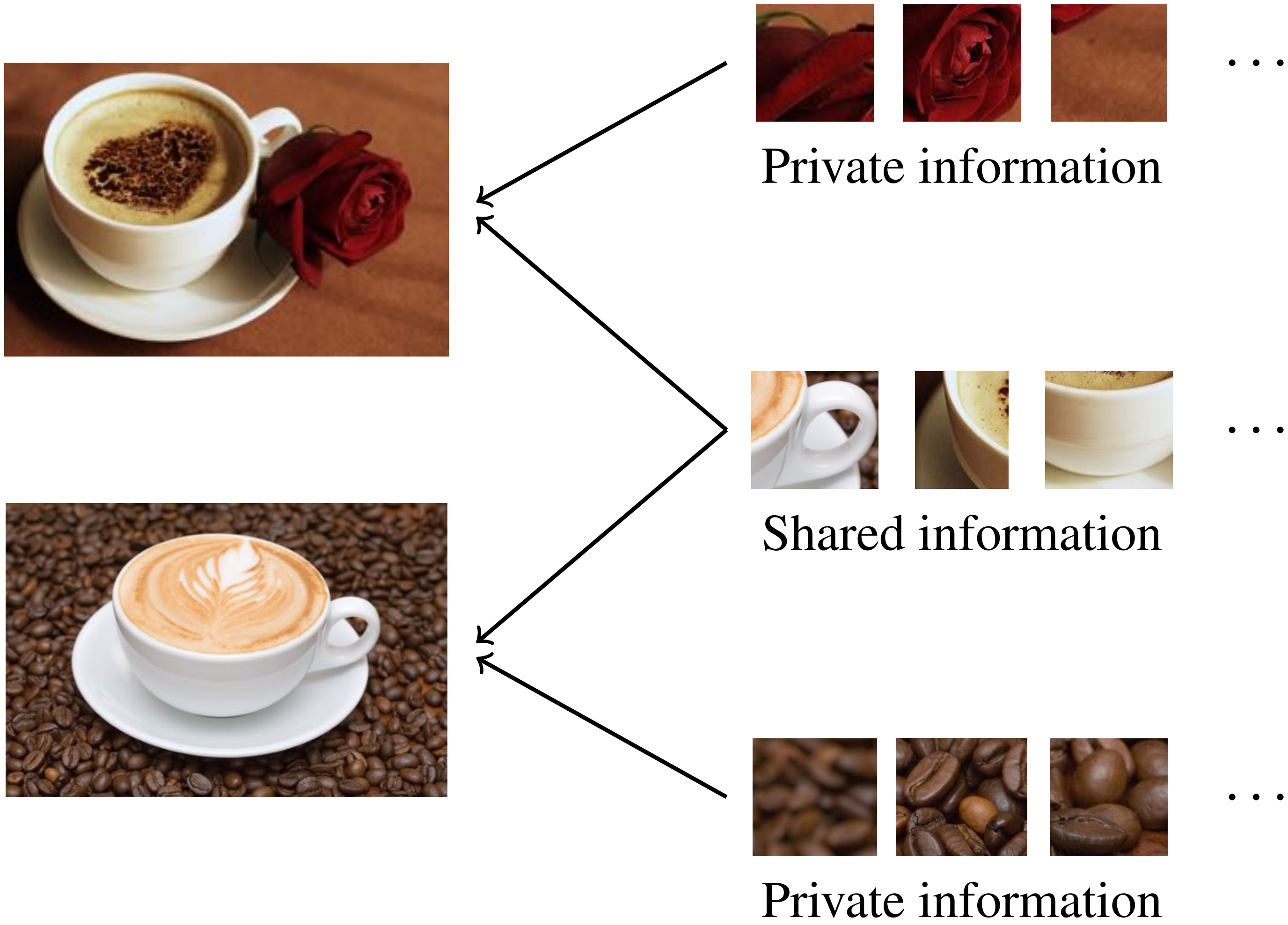


Tesla, Google, Uber, Nexar, Daimler, VW, Volvo, …

But in some cases

- High statespace complexity (causal chains etc)
- Data expensive (medical applications etc)
- Interpretability needed (financial, medical applications etc)

Structured Latent Representation – Inter-Battery Topic Model



Cheng Zhang
PhD 2016

Structured Latent Representation – Inter-Battery Topic Model

Cheng Zhang
PhD 2016

private information

…

cup **rose**

shared information

I prepared a cup of coffee with a red rose for my boyfriend.

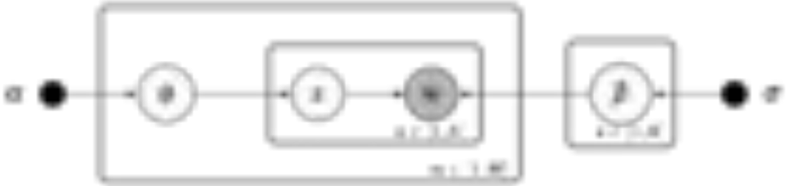
I; and; boyfriend …

private information

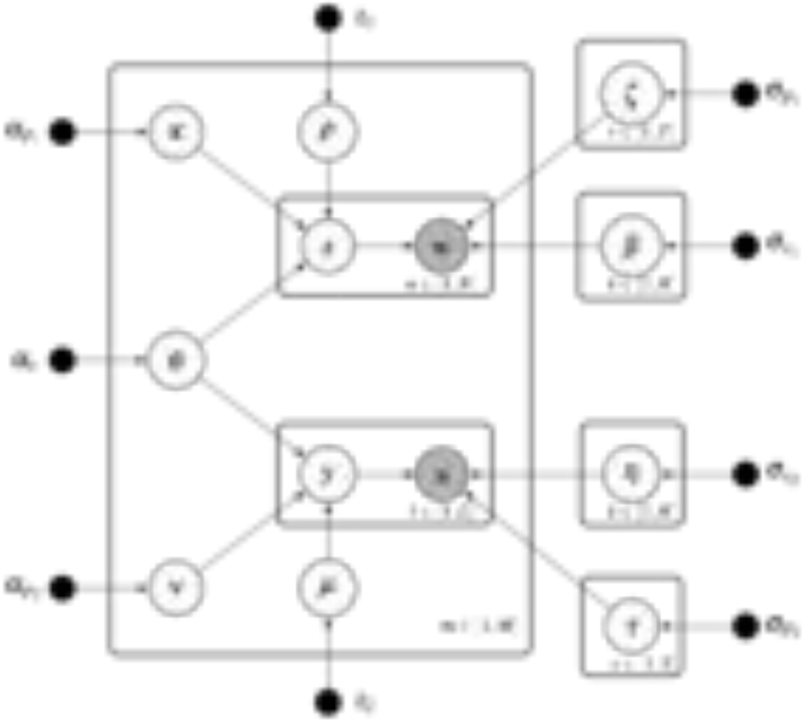
C. Zhang, H. Kjellström, and C. H. Ek. Inter-battery topic representation learning. *European Conference on Computer Vision*, 2016

Structured Latent Representation – Inter-Battery Topic Model

Cheng Zhang
PhD 2016



LDA



IBTM

$$\rho \sim Beta(\iota_1)$$

$$\mu \sim Beta(\iota_2)$$

$$z \sim Mult([\rho * \theta; (1-\rho) * \kappa])$$

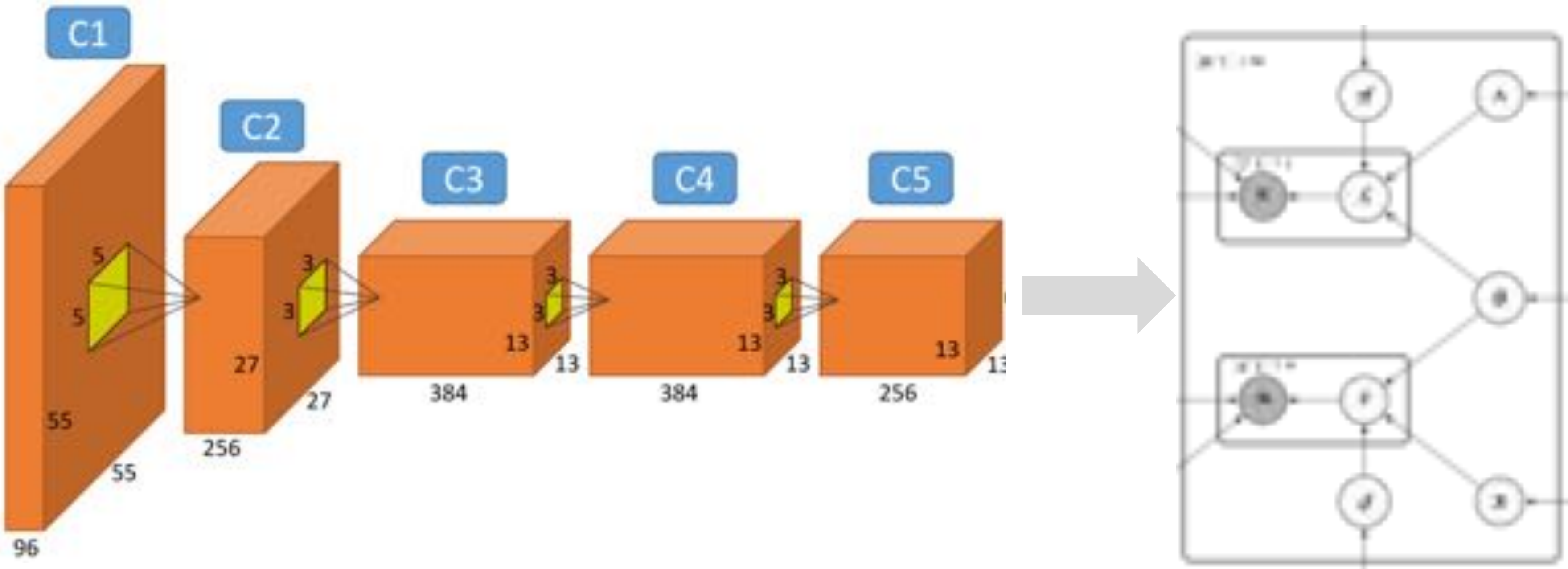
$$y \sim Mult([\mu * \theta; (1-\mu) * \nu])$$

C. Zhang, H. Kjellström, and C. H. Ek. Inter-battery topic representation learning. *European Conference on Computer Vision*, 2016

Structured Latent Representation – Inter-Battery Topic Model

Cheng Zhang
PhD 2016

Better classification results on ImageNet than a regular CNN structure



C. Zhang, H. Kjellström, and C. H. Ek. Inter-battery topic representation learning. *European Conference on Computer Vision*, 2016

Principled Combination of Logic and Deep Learning

Samuel Murray

Probabilistic framework

Deep Gaussian Processes (Neil Lawrence et al.)

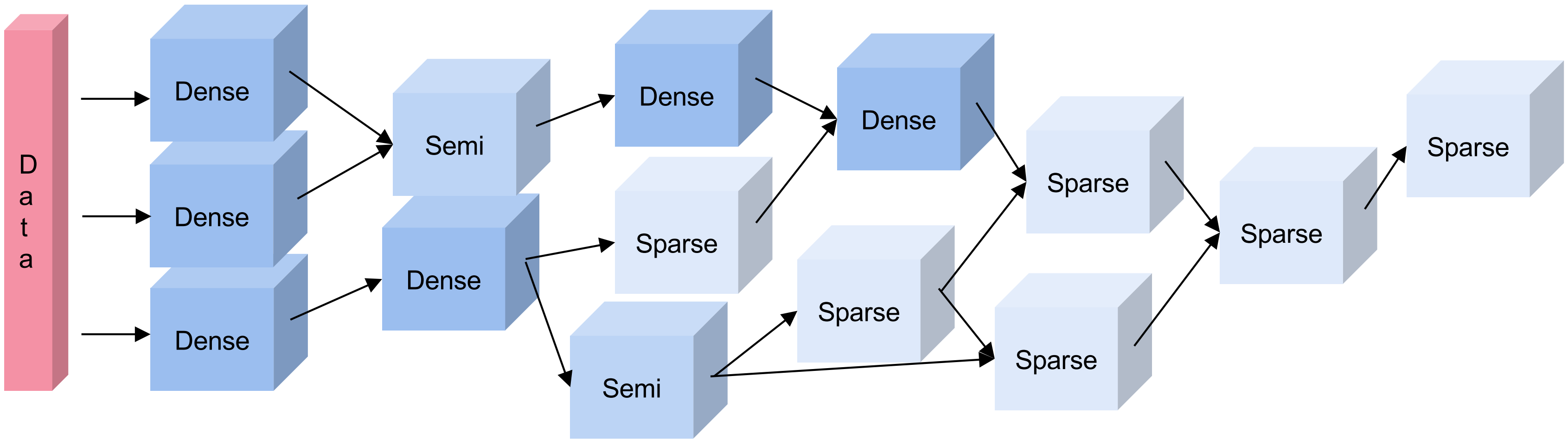
Deep Exponential Families (David Blei et al.)

Sparsity/structural prior

As simple and ”logic-like” representation as possible

Expert/Intutitive physics/intuitive psychology priors

Probabilistic framework enables principled integration of additional priors

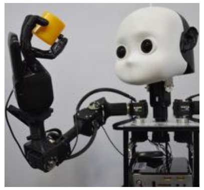


Coming soon… ☺

Conclusion

Embodiment shapes the way humans interact and learn

1. Perception-action loop



2. Low communication bandwidth

3. Learning from few examples

Take it into consideration when designing embodied artificial systems!