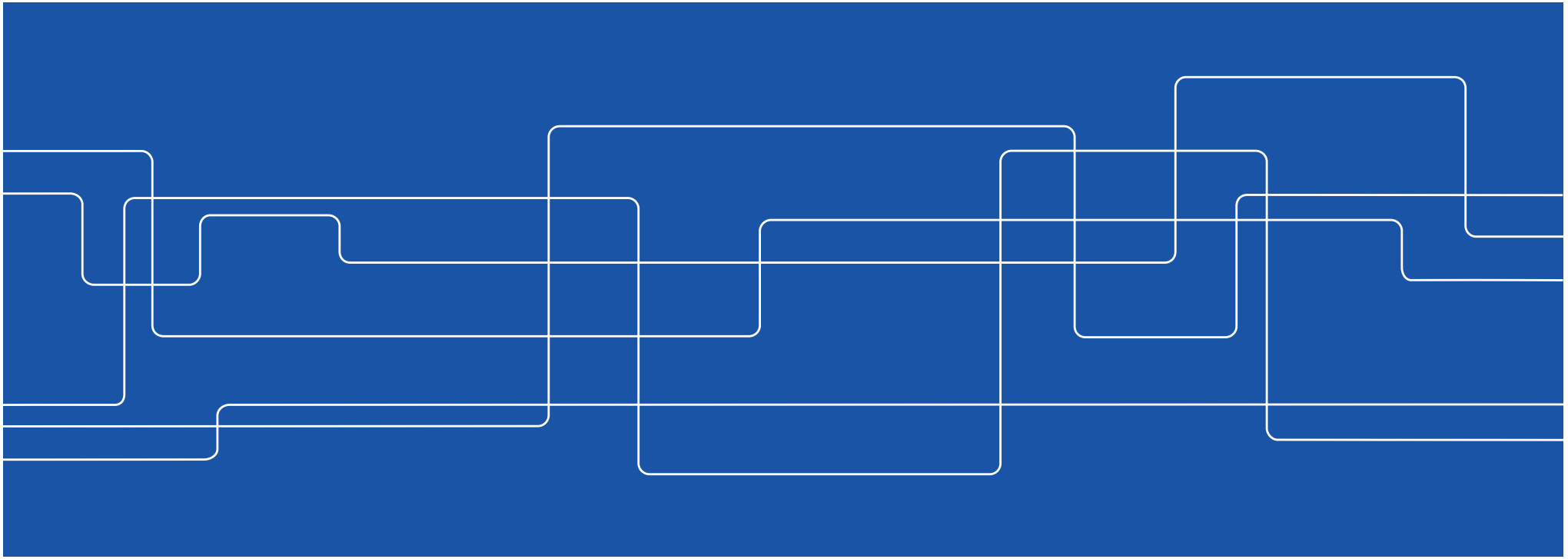
KTH ROYAL INSTITUTE
OF TECHNOLOGY

How to Make Artificial Agents a Bit More Like Us

Hedvig Kjellström

Professor of Computer Science

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



Why make artificial agents that function like humans?

To function in a world made for humans, agents need to:

1. Interact with humans

2. Learn online and from few examples like humans

Embodiment – key to what is human-like

What is Embodiment?

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

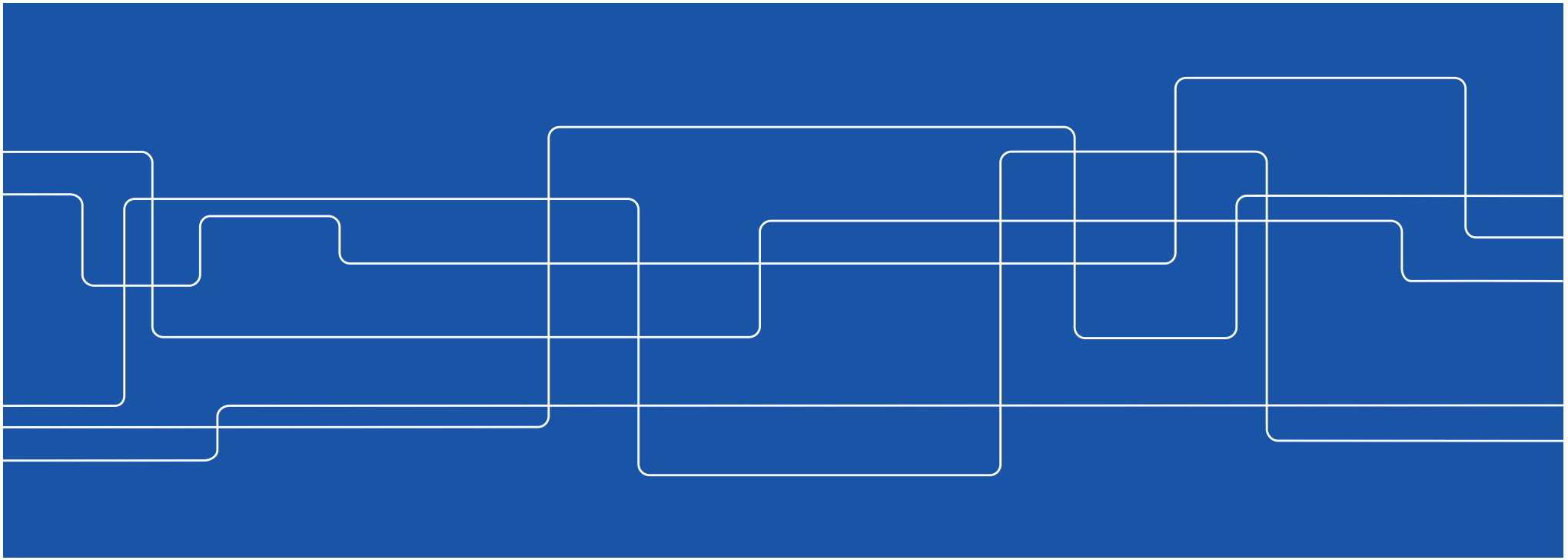
How does it affect the way we function?

Studied in the field of Embodied Cognition

1. **Interact**
2. **Learn**

M. V. Butz and E. F. Kutter. *How the Mind Comes into Being*, 2017
R. Pfeifer and J. Bongard. *How the Body Shapes the Way We Think*, 2007

Aspect 1: Interact



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: $E \approx 10^{16}$

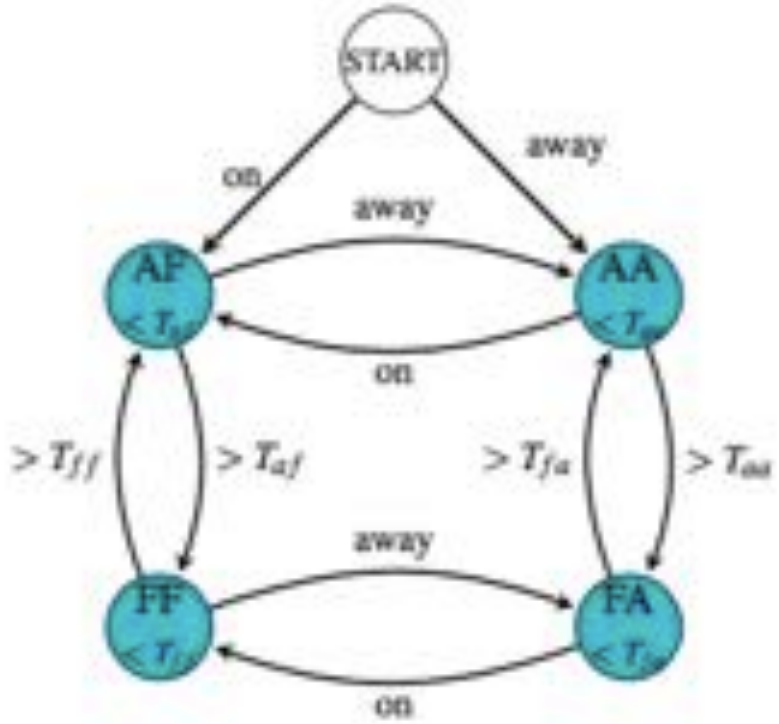
Computer: $E \approx 10$

Conclusions

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

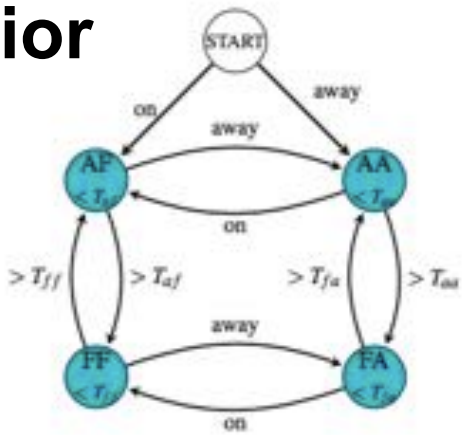
Perception and Production of Gaze Aversion Behavior

Yanxia Zhang
PostDoc 2016

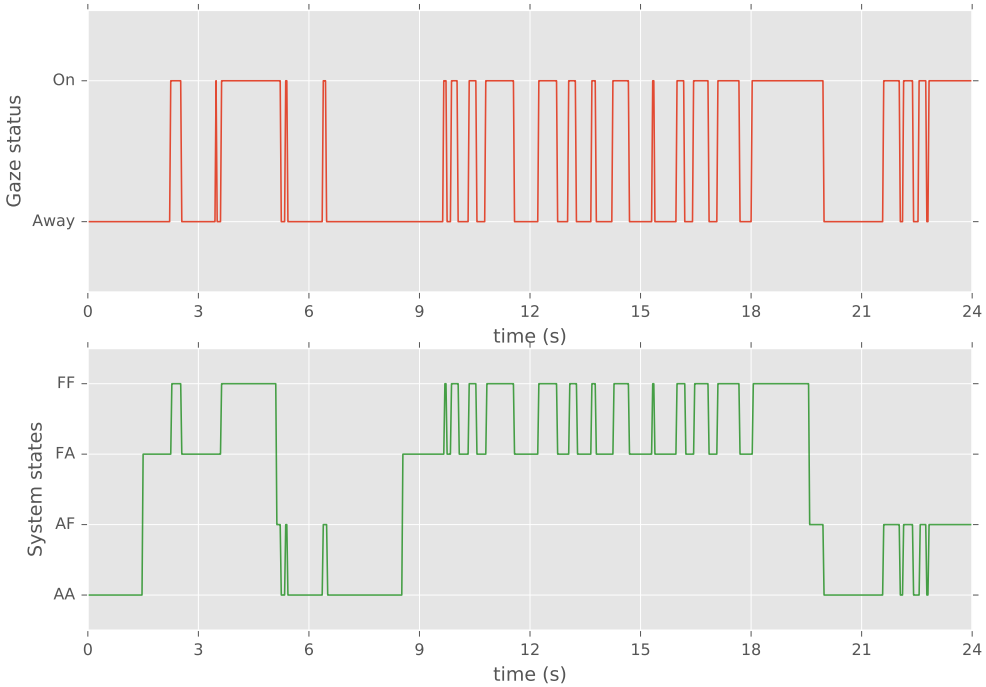


Y. Zhang, J. Beskow, and H. Kjellström. Look but don't stare: Mutual gaze interaction in social robots. *International Conference on Social Robotics*, 2017

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

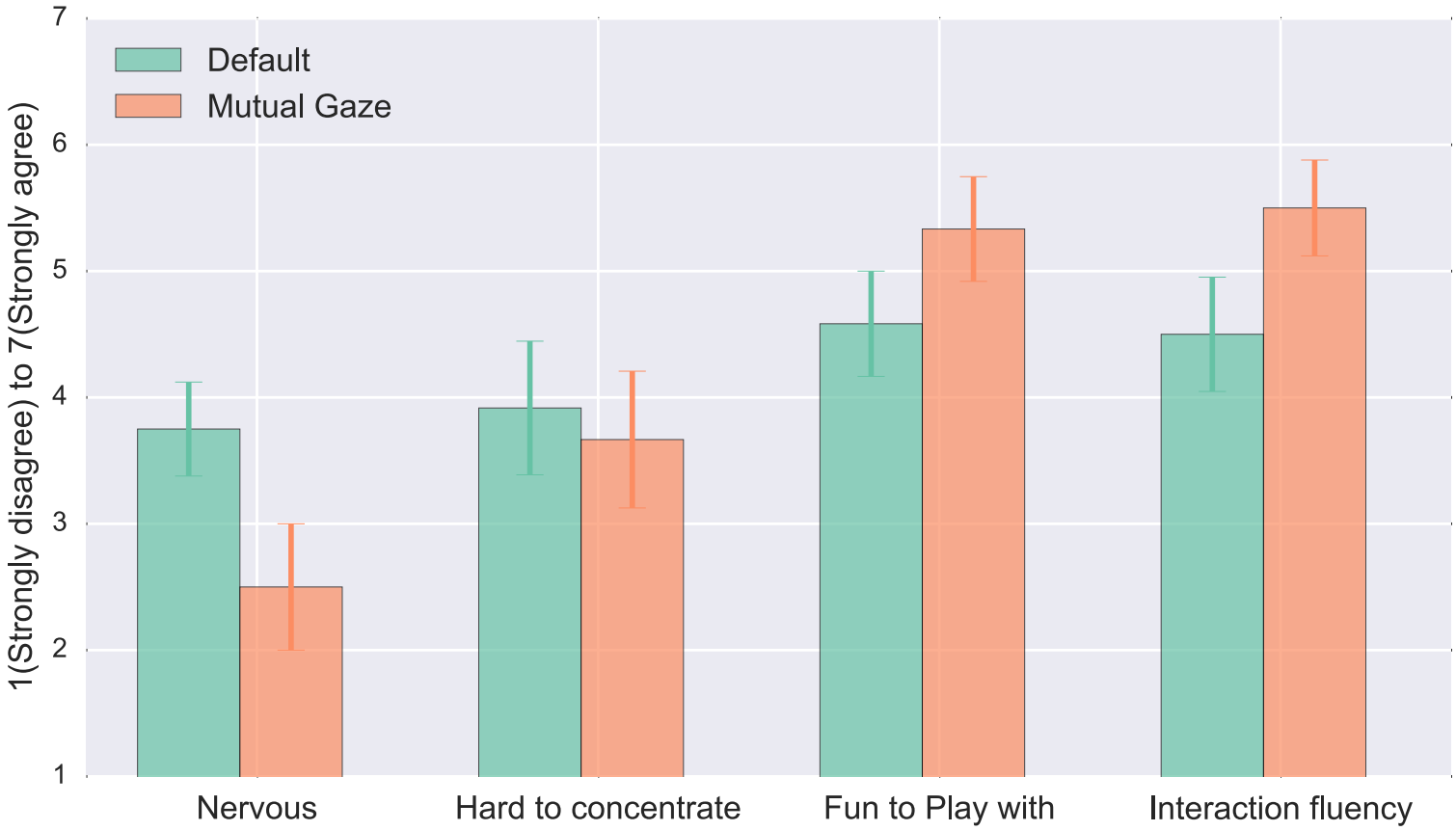
Perception and Production of Gaze Aversion Behavior

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

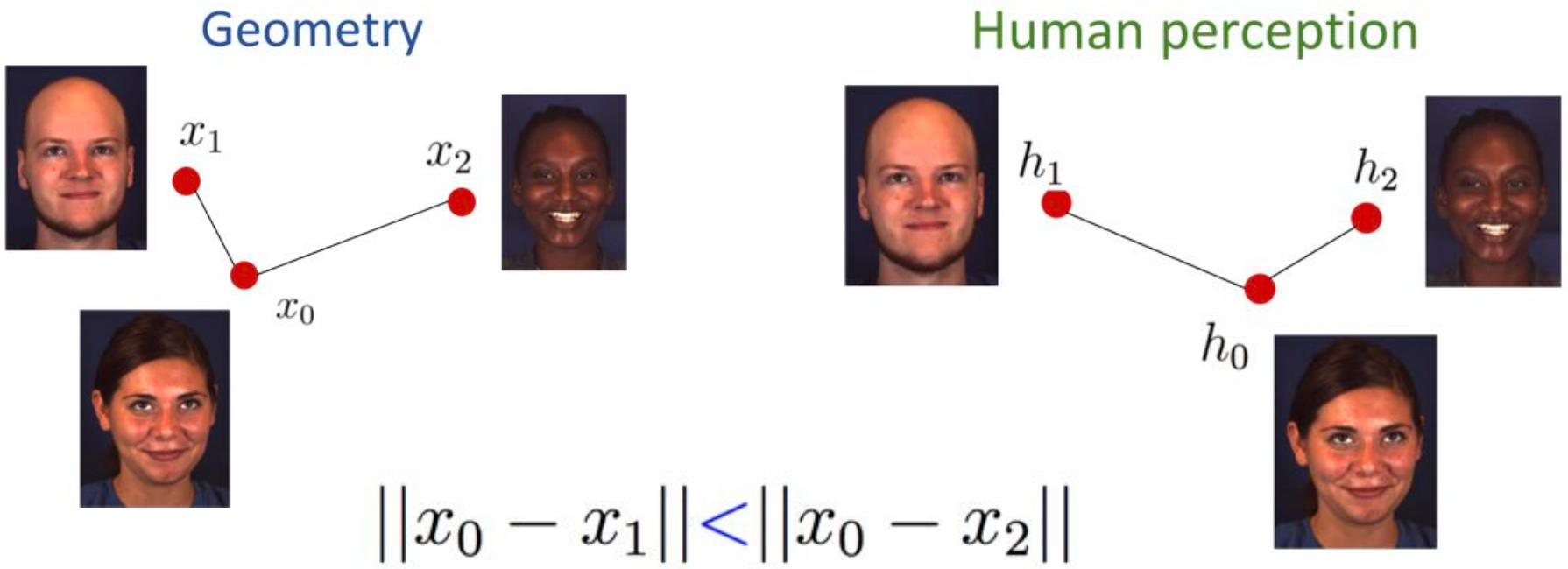
Yanxia Zhang
PostDoc 2016



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

Human-Like Perception of Facial Expression

Olga Mikheeva
PhD student

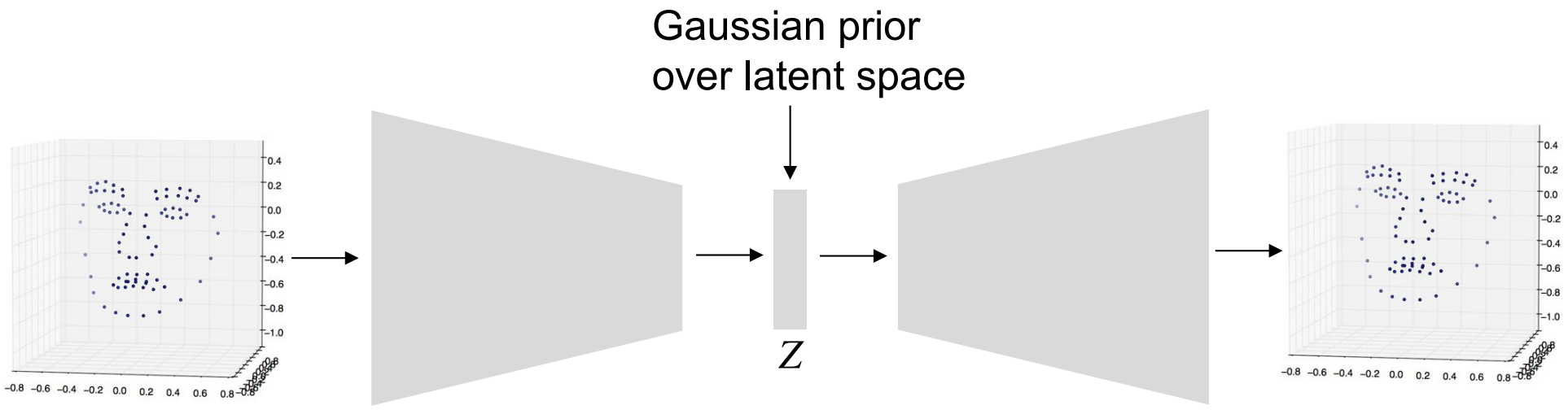


O. Mikheeva, C. H. Ek, and H. Kjellström. Perceptual facial expression representation. *International Conference on Automatic Face and Gesture Recognition*, 2018

Human-Like Perception of Facial Expression

Olga Mikheeva
PhD student

Standard VAE with Gaussian prior

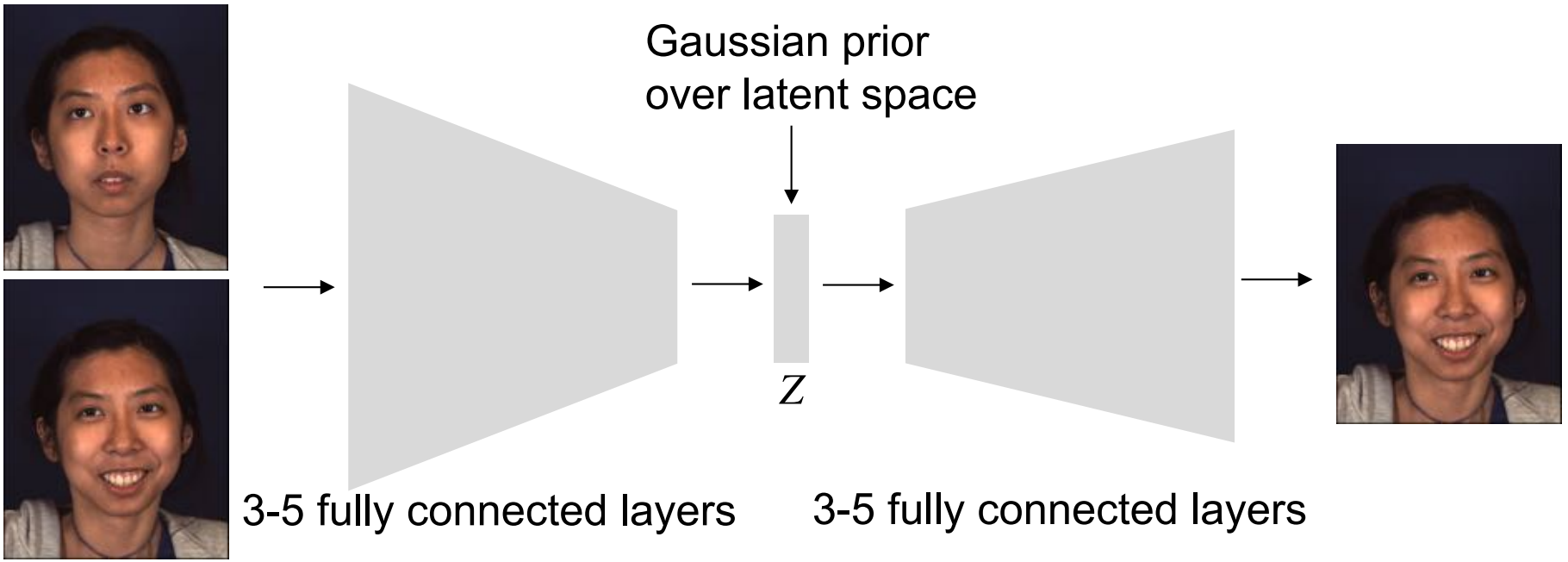


O. Mikheeva, C. H. Ek, and H. Kjellström. Perceptual facial expression representation. *International Conference on Automatic Face and Gesture Recognition*, 2018

Human-Like Perception of Facial Expression

Olga Mikheeva
PhD student

Model *M1*, VAE with **neutral face**

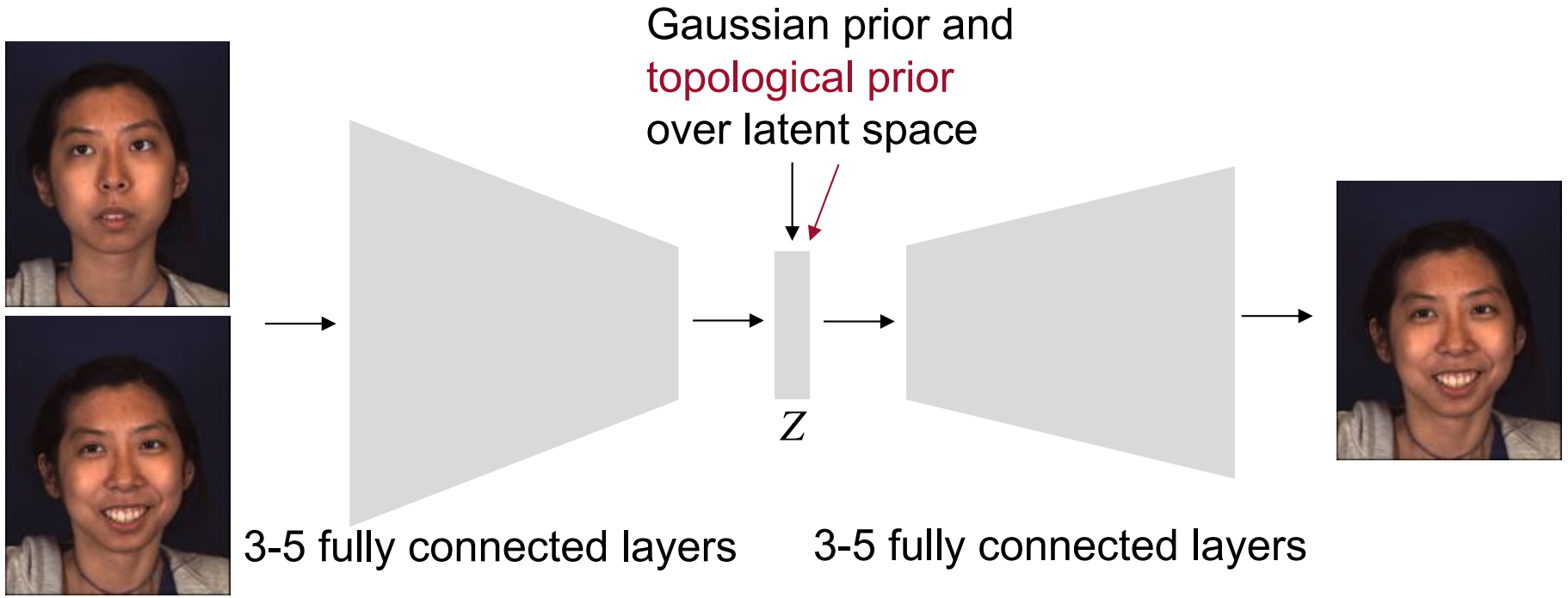


O. Mikheeva, C. H. Ek, and H. Kjellström. Perceptual facial expression representation. *International Conference on Automatic Face and Gesture Recognition*, 2018

Human-Like Perception of Facial Expression

Olga Mikheeva
PhD student

Model *M2*, VAE with neutral face and **topological prior**



O. Mikheeva, C. H. Ek, and H. Kjellström. Perceptual facial expression representation. *International Conference on Automatic Face and Gesture Recognition*, 2018

Human-Like Perception of Facial Expression

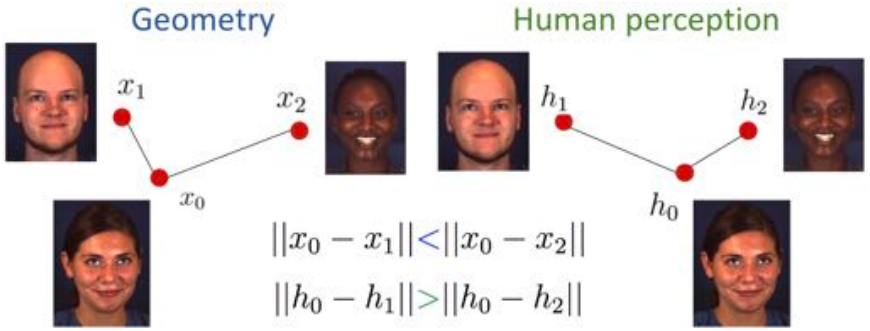
Olga Mikheeva
PhD student

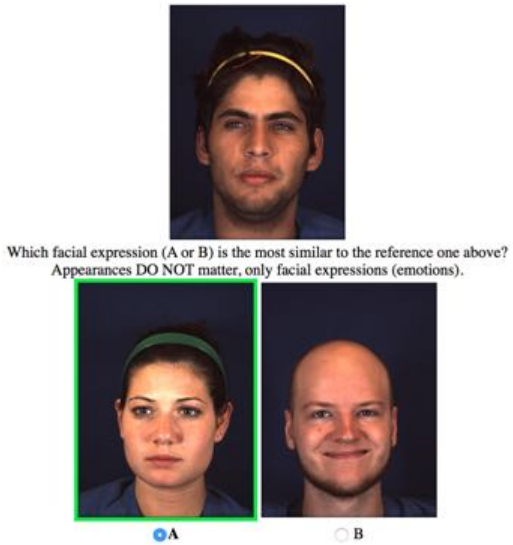
Topological prior

Penalize incoherency with human perception

Human perception triplets

$$\mathbf{S} = \{(s_t^{ref}, s_t^+, s_t^-), w_t\}_{t=1}^T$$





$$p_T(\mathbf{Z}|\mathbf{S}) \propto \exp(-\frac{1}{\gamma}\Phi(\mathbf{Z},\mathbf{S})) \quad \textit{where}$$

$$\Phi(\mathbf{Z},\mathbf{S}) = \sum_{i=1}^{T} max\big(0; d(\mathbf{z}^{(s_t^{ref})}, \mathbf{z}^{(s_t^+)}) - d(\mathbf{z}^{(s_t^{ref})}, \mathbf{z}^{(s_t^-)})\big)$$

For BU-3DFE (3D static posed) human triplets generated from expression labeling

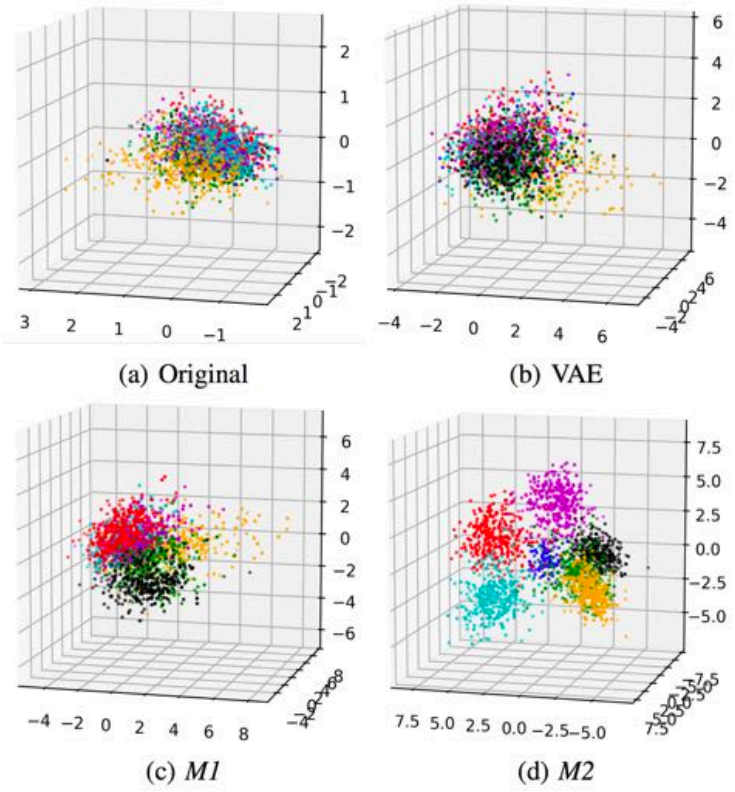
For BP-4DSFE (3D dynamic spontaneous) human triplets collected using crowdsourcing

O. Mikheeva, C. H. Ek, and H. Kjellström. Perceptual facial expression representation. *International Conference on Automatic Face and Gesture Recognition*, 2018

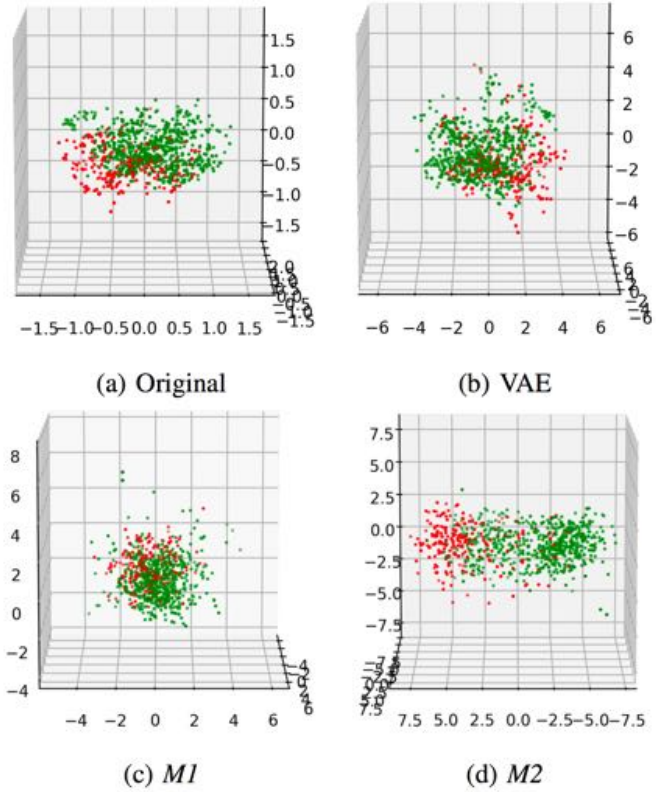
Human-Like Perception of Facial Expression

Olga Mikheeva
PhD student

Latent space (3 principal components)



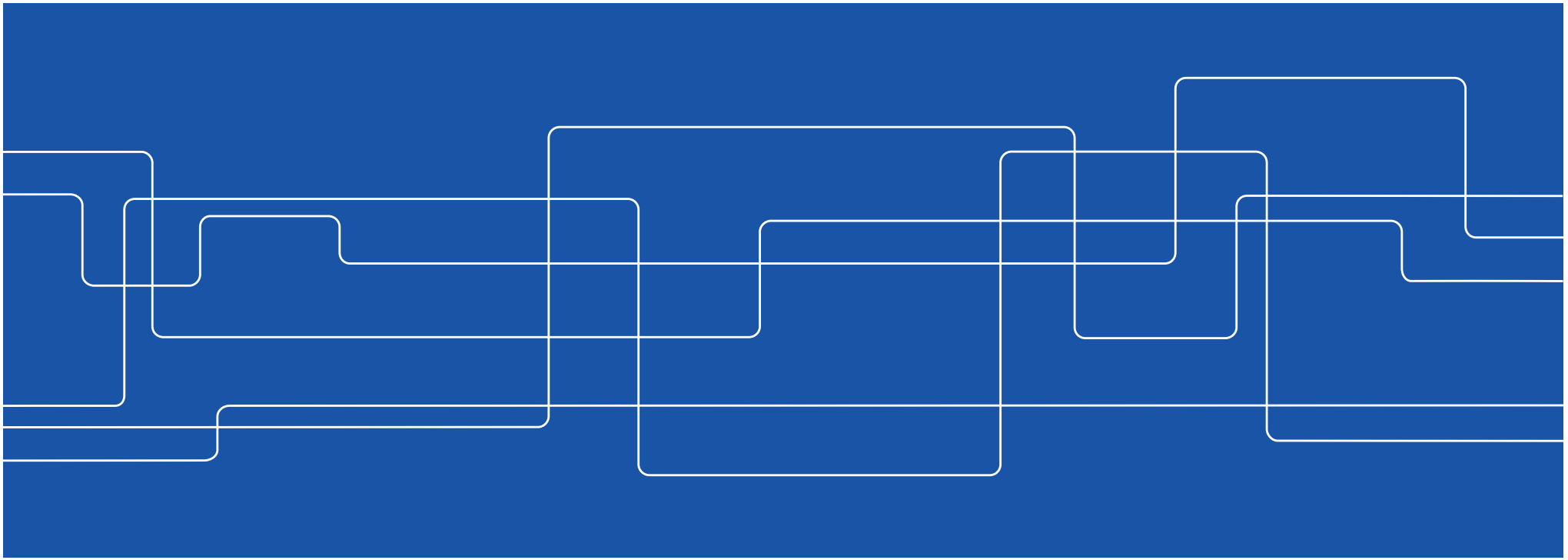
Static, posed dataset
(angry/disgusted/sad/afraid/surprised/happy/neutral)



Dynamic, spontaneous dataset
(positive/negative)

O. Mikheeva, C. H. Ek, and H. Kjellström. Perceptual facial expression representation. *International Conference on Automatic Face and Gesture Recognition*, 2018

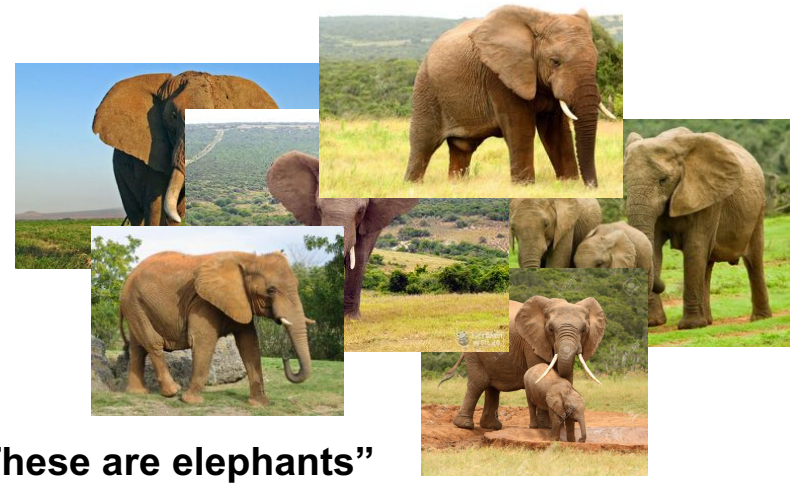
Aspect 2: Learn



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

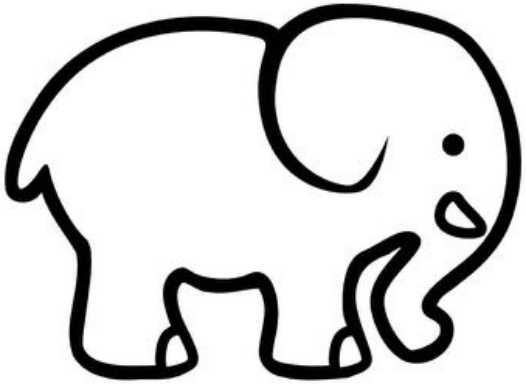
Embodiment Shapes the Way We Learn – Learning from Few Examples

State of the art ML algorithm



”These are elephants”

Toddler



”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

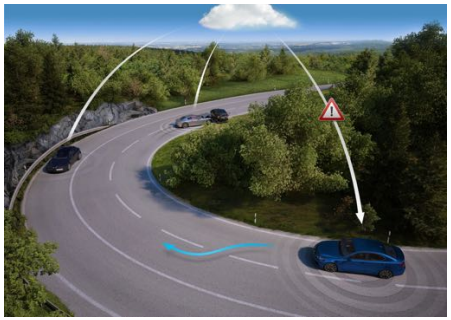
Embodiment Shapes the Way We Learn – But Still Learn from Many Examples?

Alternative strategy – provide enough training data!

Crowd Sourcing

Hey there! I'm a robot brain. I learn concepts by searching the Internet. I can interpret natural language text, images, and videos. I watch humans with my sensors and learn things from interacting with them. Here are a few things I've learned recently...

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

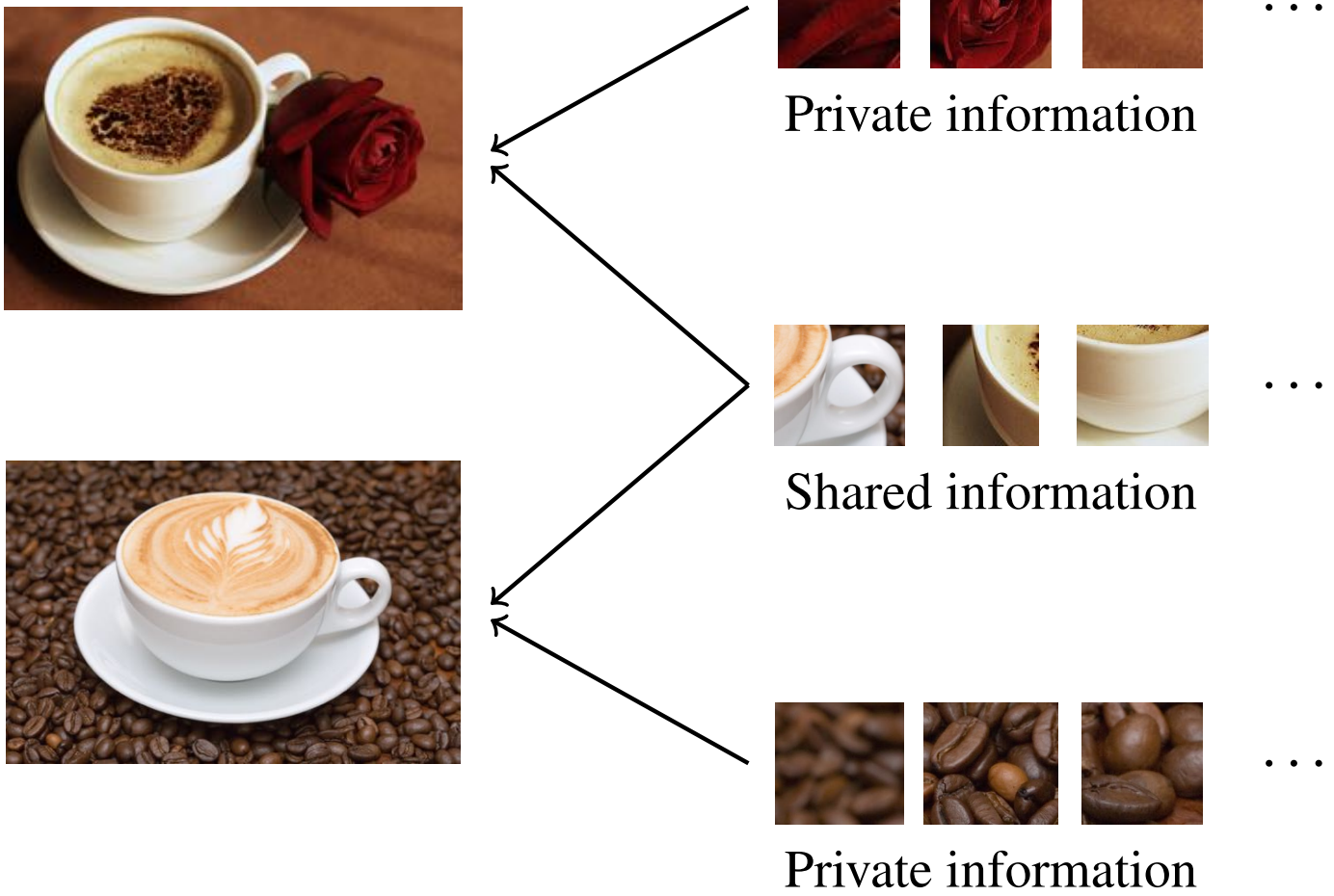


Tesla, Google, Uber, Nexar, Daimler, VW, Volvo, …

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



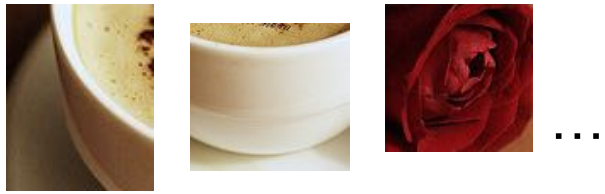
Cheng Zhang
PhD 2016

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

private information



cup **rose**

shared information

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

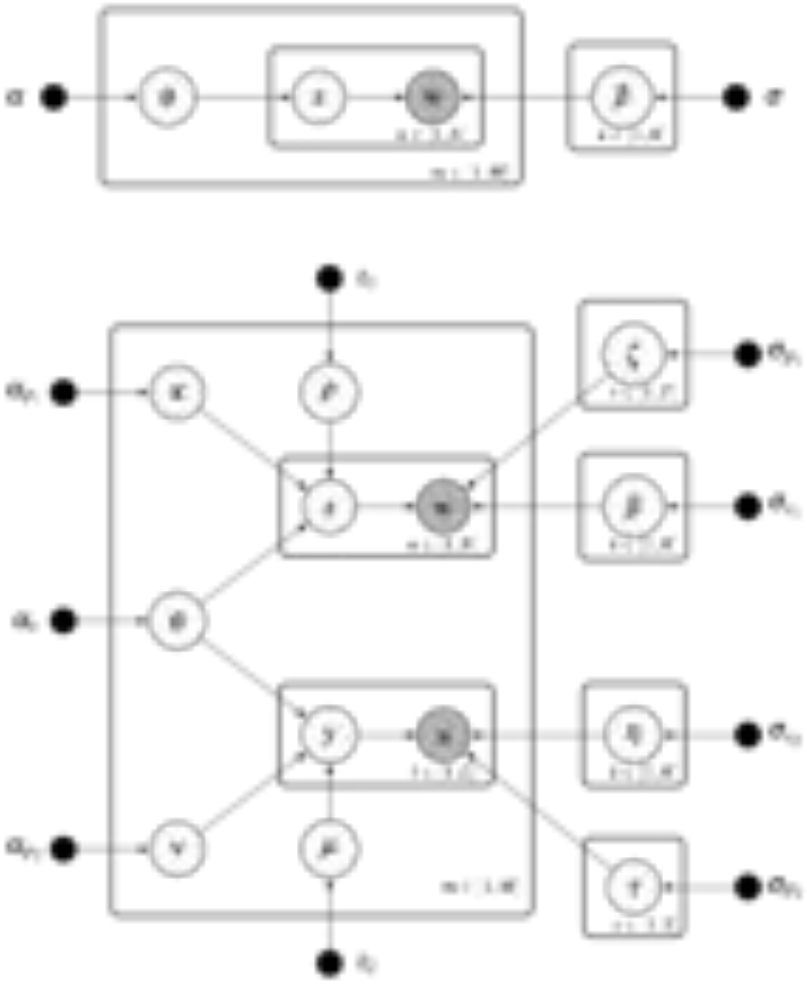
I; and; boyfriend …

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



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

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

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

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

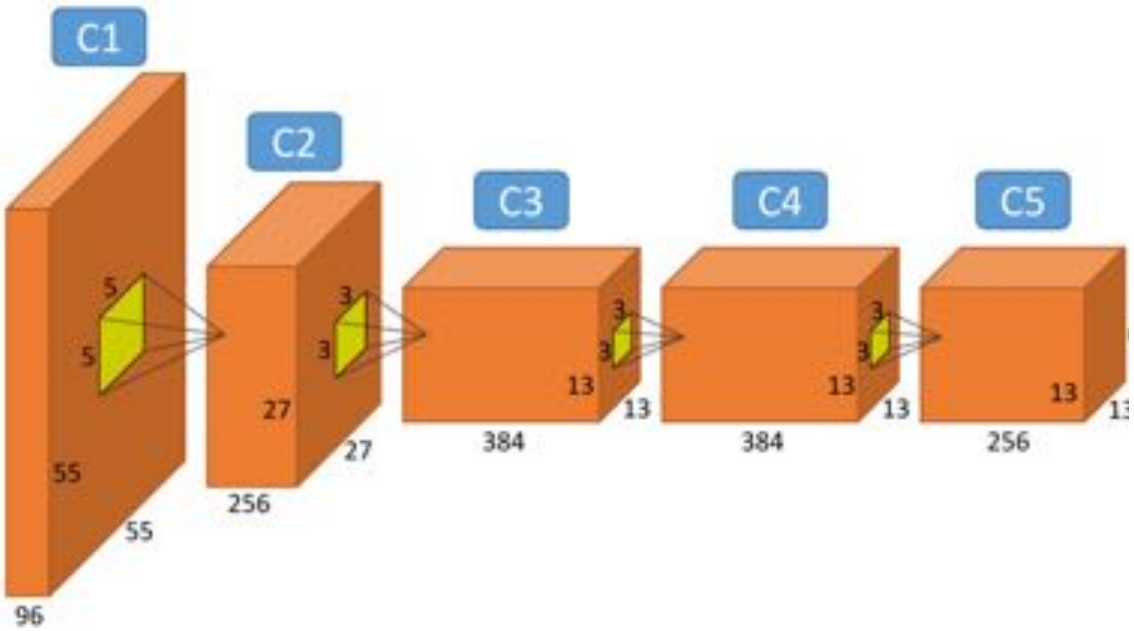
C. Zhang, H. Kjellström, and C. H. Ek. Inter-battery topic representation learning. *European Conference on Computer Vision*, 2016

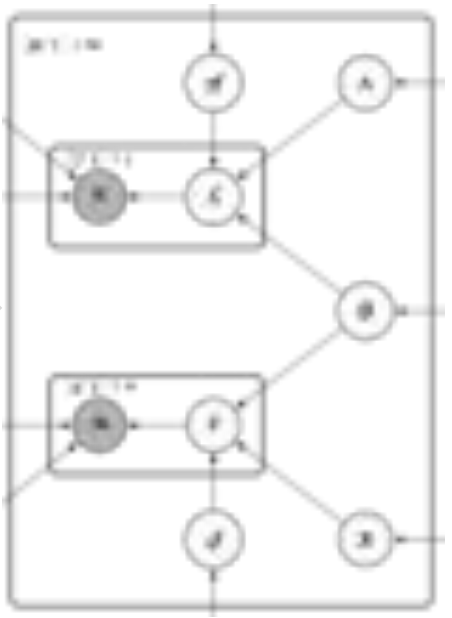
Structured Latent Representation – Inter-Battery Topic Model

Cheng Zhang
PhD 2016

CNN close to data, PGM higher up

Better classification results on ImageNet than a regular CNN structure





Conclusion

Artificial agents should be made human-like

The essence of human-like: embodiment, shapes the way humans interact and learn

1. Low communication bandwidth

2. Learning from few examples

Take it into consideration when designing embodied artificial systems!

Thanks to my Collaborators!


Taras Kucherenko


Marcus Klasson


Olga Mikheeva


Sofia Broomé


Samuel Murray


Ruibo Tu


Judith Bütepage
Joint with Danica Kragic


Cheng Zhang
Microsoft Research
Cambridge, UK


Yanxia Zhang
TU Delft, Netherlands


Jonas Beskow
KTH Royal Institute of
Technology, Sweden


Carl Henrik Ek
University of Bristol, UK