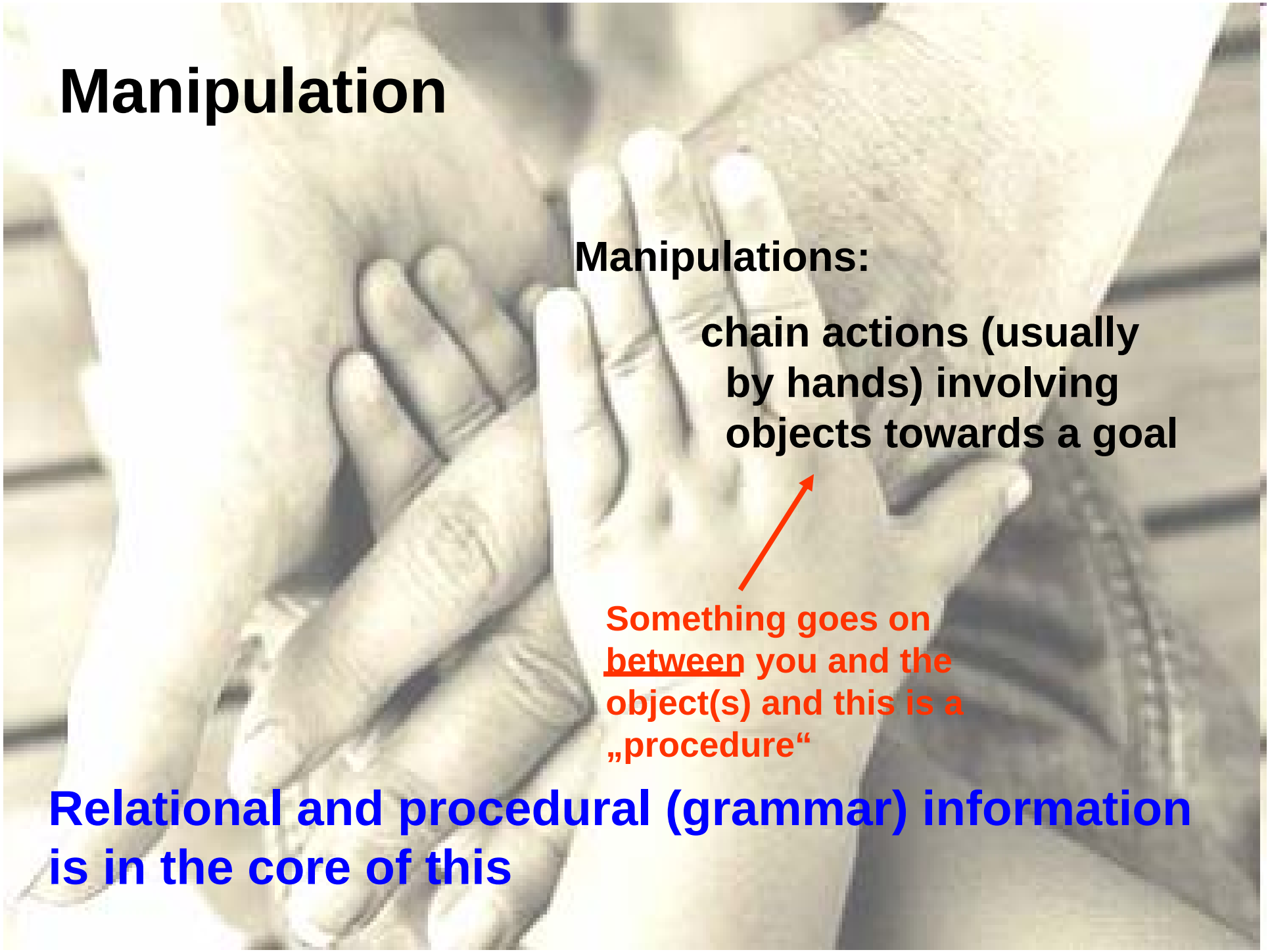
Manipulation



Manipulations:

chain actions (usually by hands) involving objects towards a goal

Something goes on <u>between</u> you and the object(s) and this is a „procedure“

Relational and procedural (grammar) information is in the core of this

The Problem: Signal to Symbol Gap (Norbert Krüger)

Recognize/“Understand“

From Pixels to AI → forming „concepts“ (and plans)

Manipulate

From AI to voltages → executing „concepts“

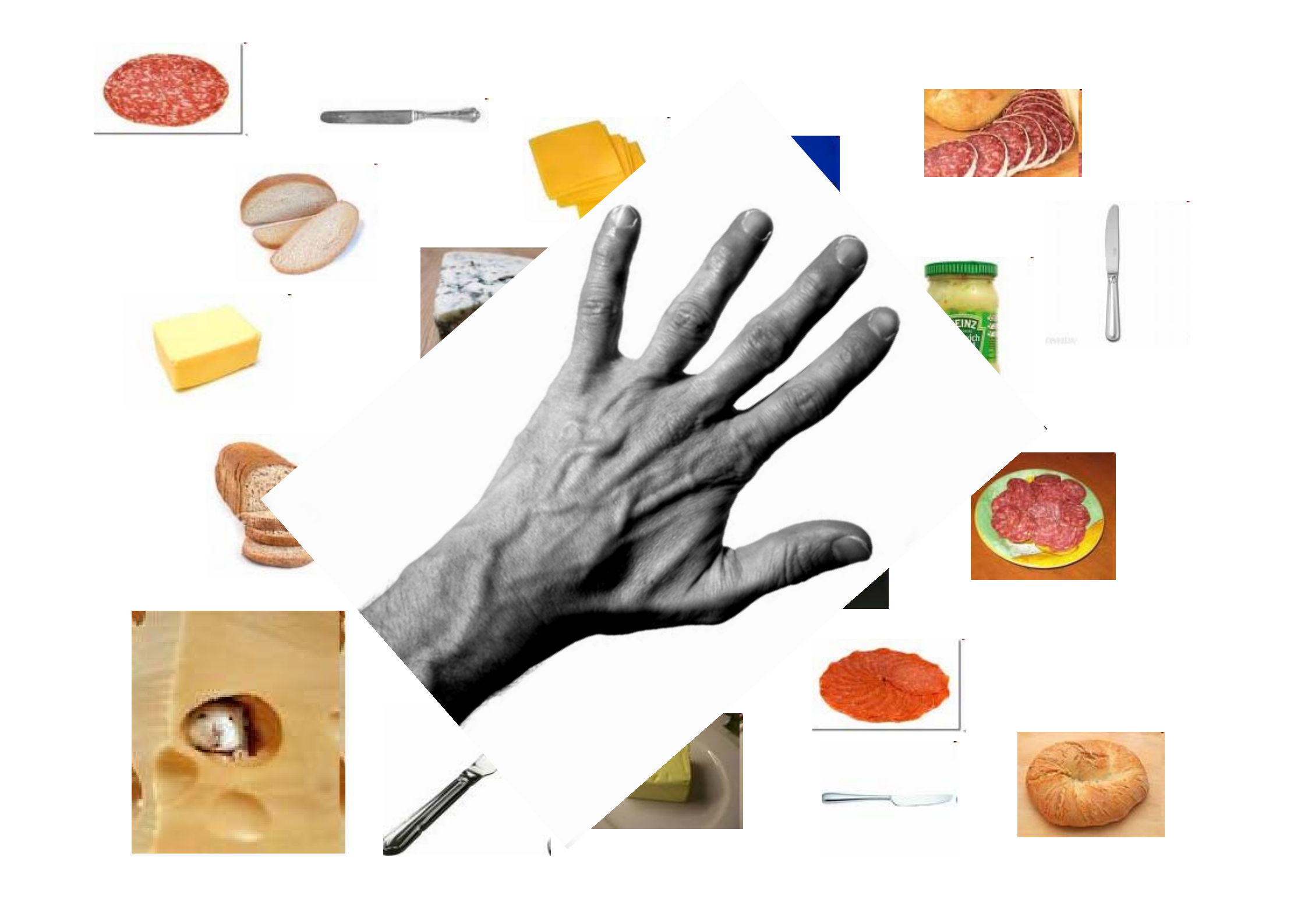
Not much discussed further

Core Question: How can such processes be bootstrapped arriving at agent-own (internal) concepts free from external models (grounding problem)

The Problem: Signal to Symbol Gap (Norbert Krüger)

Example: Understand how to „Make a sandwich“

Objects involved



The Problem: Signal to Symbol Gap (Norbert Krüger)

Example part 1: Understand how to „Make a sandwich“

Objects involved + Action

Breads

Hand Spreads

Tools

Cheese or Salami

The Problem: Signal to Symbol Gap (Norbert Krüger)

Example: Understand how to „Make a sandwich“

Objects involved + Action

Breads

Hand Spreads

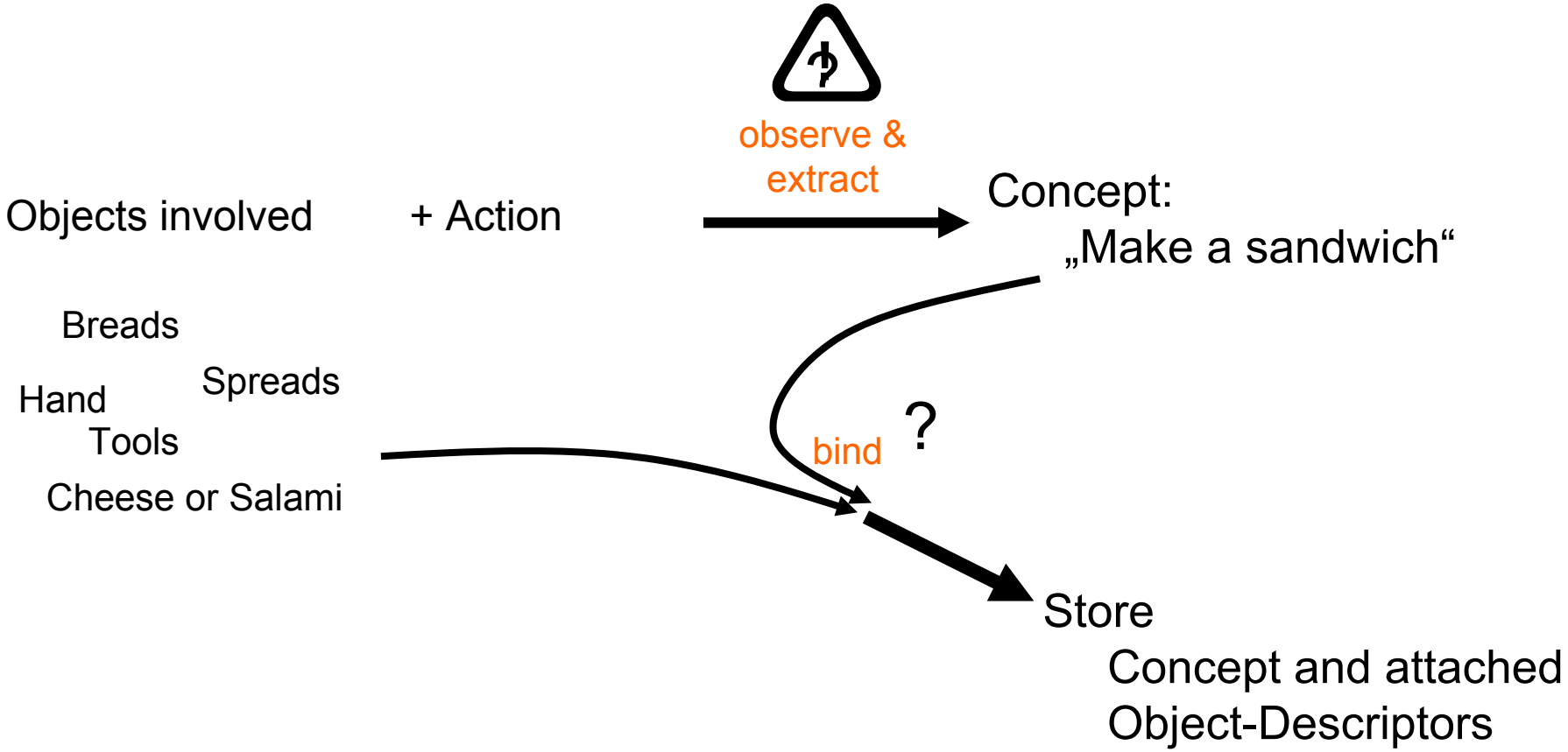
Tools

Cheese or Salami

And there is the invariance problem, too

The Problem: Signal to Symbol Gap (Norbert Krüger)

Example: Understand how to „Make a sandwich“



Core Idea:

Analyse the change of the relation between objects that occurs during a manipulation.

Needed:

Object descriptors → Segments
Tracking
Depth

Manipulation descriptors (grammar)
Object categorization WITHIN a manipulation
Learning

Core Idea: **Object Action Complexes**

Analyse the change of the relation between objects that occurs during a manipulation.

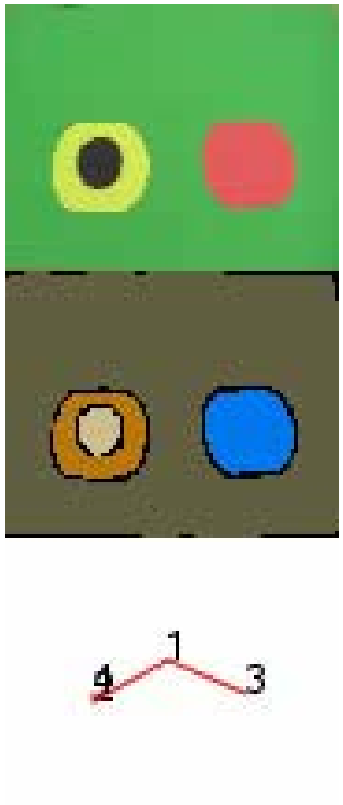
Categorization WITHIN a manipulation context.

Algorithm – Artificial Scenario

Original Images

Segementation & Tracking

Continuous Graphs

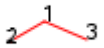


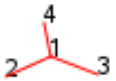
Note the deleting and forming of edges

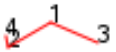
Make linked graphs of segment centers

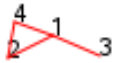
Extract and track all segment centers

Algorithm – Object Relations

$\rho_{4,2}$  = **Absence** = **-1**

$\rho_{4,2}$  = **No Connection** = **0**

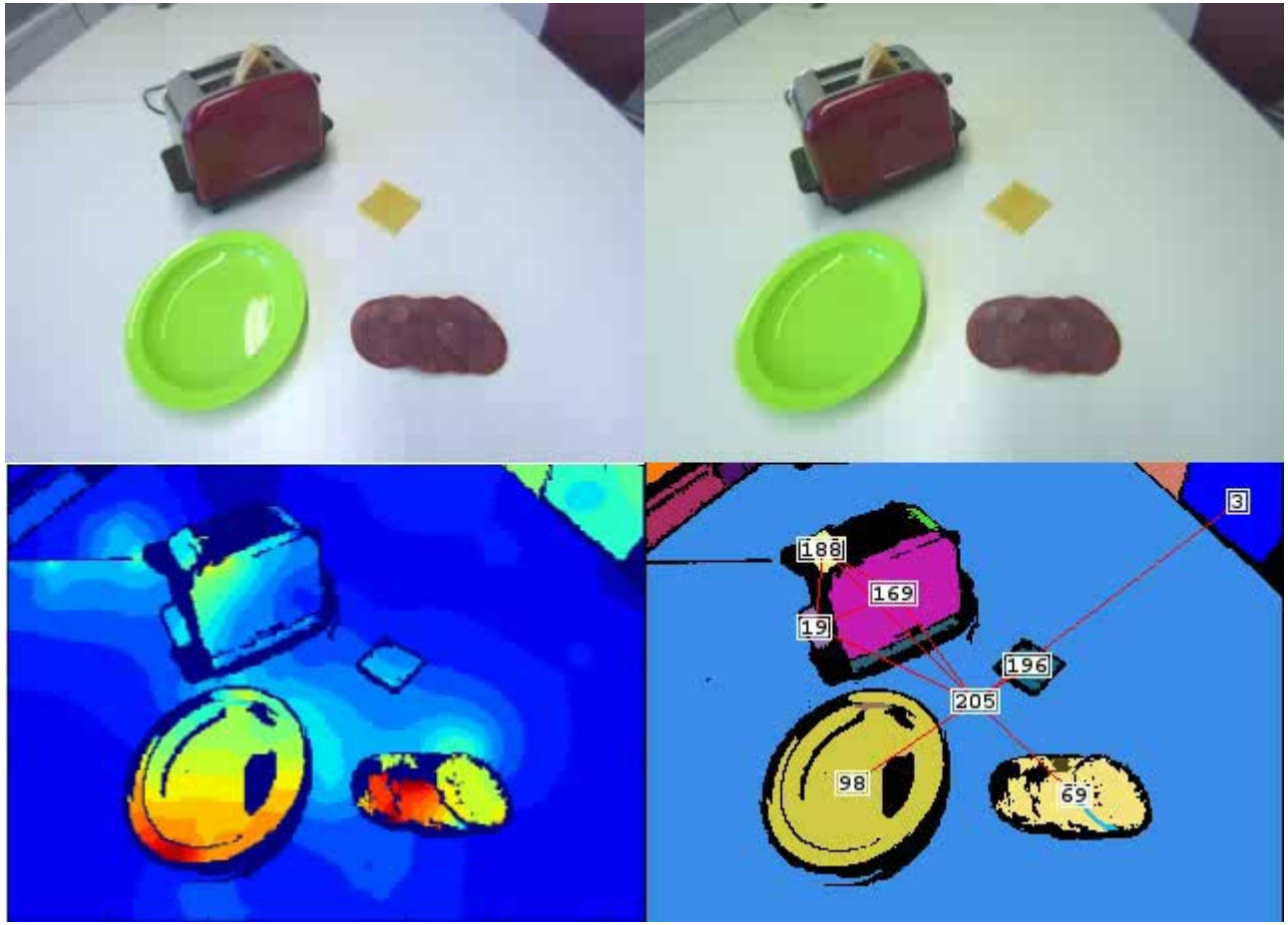
$\rho_{4,2}$  = **Overlapping** = **1**

$\rho_{4,2}$  = **Touching** = **2**

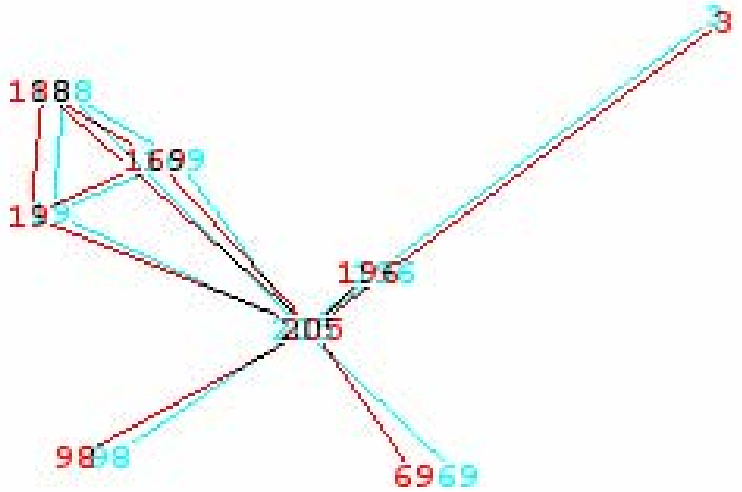
$\rho_{4,2}$ } **Object #2 :** *Full-to-Empty (yellow) Vessel*
Object #4 : *Content*

Reducing Complexity and adding Object Models

3D-Tracked Graphs in Real Time

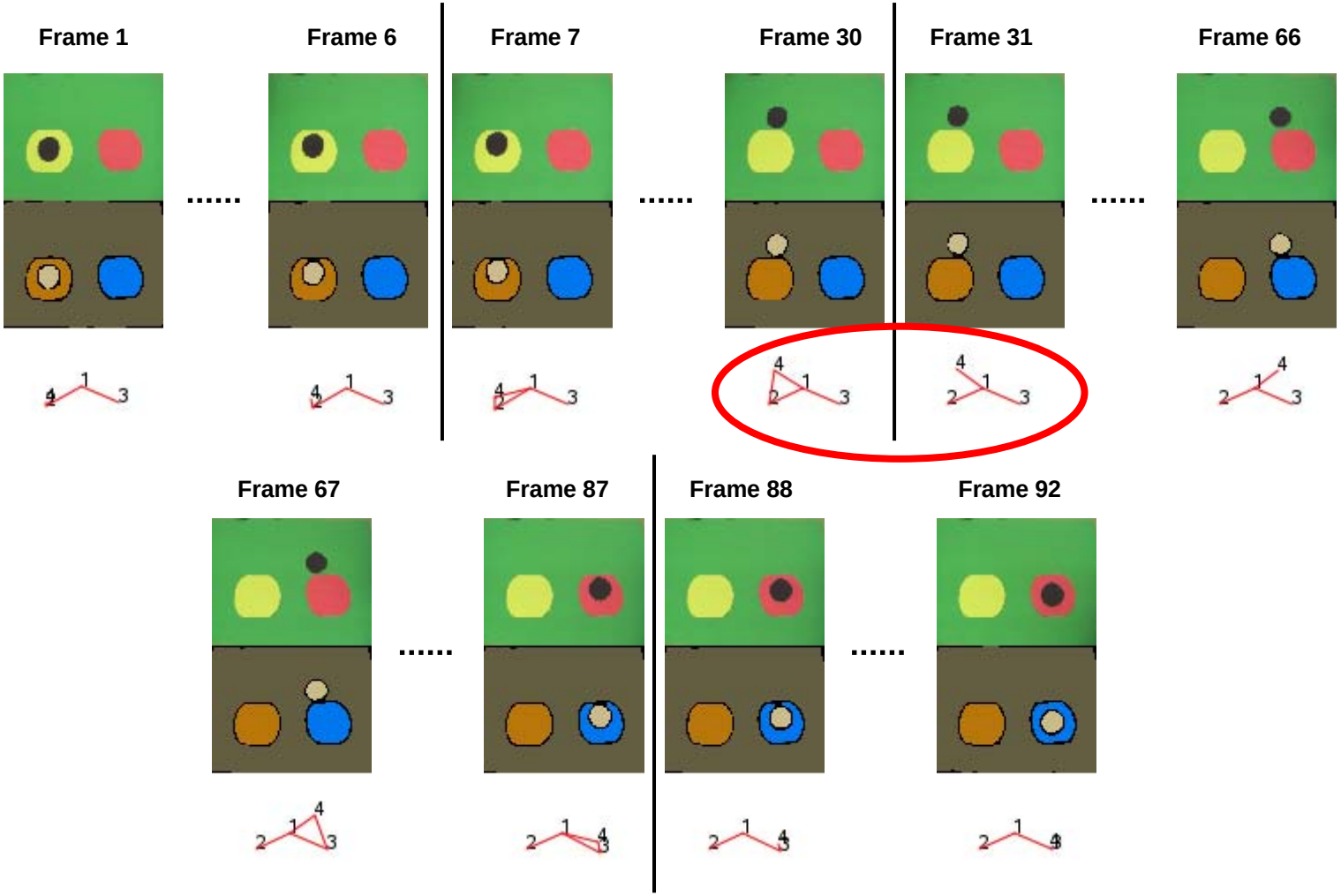


Anaglyph
Representation



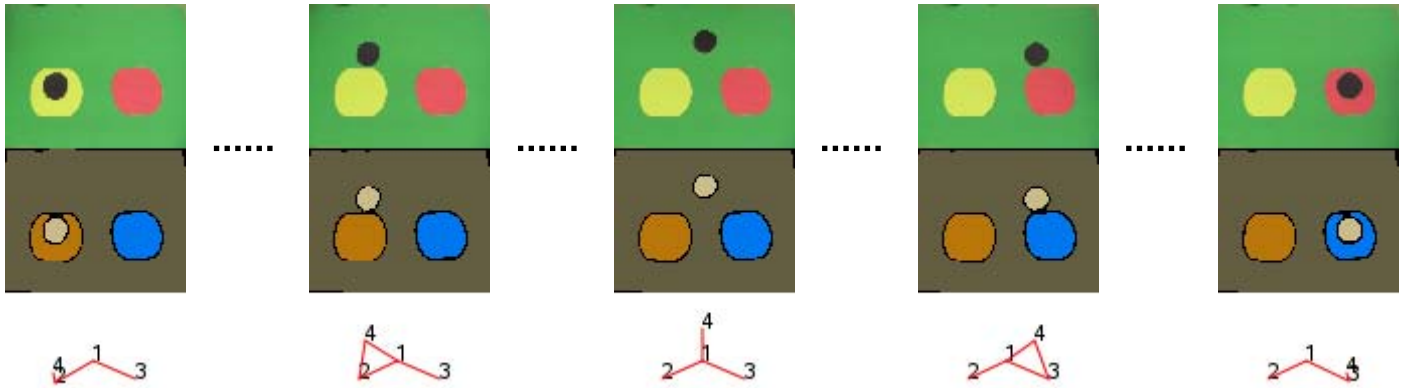
Extract and track all segment centers

Algorithm – The essence is to extract the Topological Changes in the linked graphs as these are indicative of a “Change”



Extract and track all segment centers

Algorithm – Event Table



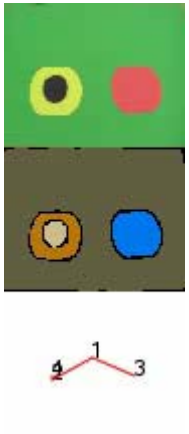
$\rho_{2,1}$	2	2	1	1	1
$\rho_{3,1}$	1	1	1	2	2
$\rho_{4,1}$	0	2	1	2	0
$\rho_{4,2}$	1	2	0	0	0
$\rho_{4,3}$	0	0	0	2	1

5X5

This table fully represents a manipulation

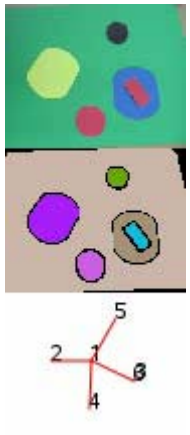
Compare to other „manipulation“

Algorithm – Similarity Measure



$$\begin{matrix}\rho_{2,1}\\ \rho_{3,1}\\ \rho_{4,1}\\ \rho_{4,2}\\ \rho_{4,3}\end{matrix}\begin{pmatrix}2 & 2 & 1 & 1 & 1\\ 1 & 1 & 1 & 2 & 2\\ 0 & 2 & 1 & 2 & 0\\ 1 & 2 & 0 & 0 & 0\\ 0 & 0 & 0 & 2 & 1\end{pmatrix}$$

$$\begin{matrix}\rho_{2,1}\\ \rho_{3,1}\\ \rho_{6,1}\\ \rho_{3,2}\\ \rho_{3,6}\end{matrix}\begin{pmatrix}1 & 1 & 1 & 2 & 2\\ 0 & 2 & 1 & 2 & 0\\ 2 & 2 & 1 & 1 & 1\\ 0 & 0 & 0 & 2 & 1\\ 1 & 2 & 0 & 0 & 0\end{pmatrix}$$

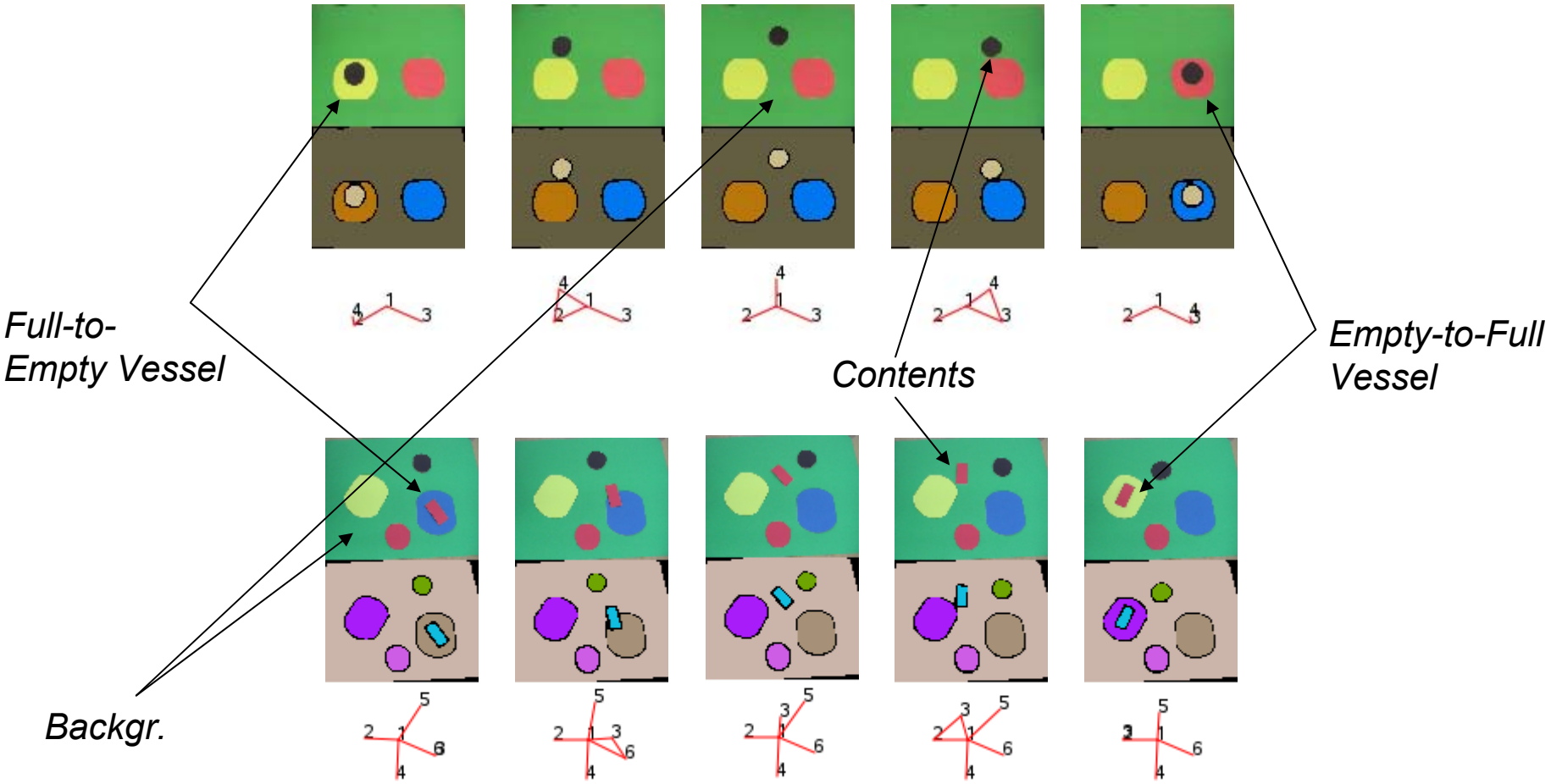


$$\begin{matrix}\rho_{2,1}\\ \rho_{3,1}\\ \rho_{4,1}\\ \rho_{4,2}\\ \rho_{4,3}\end{matrix}\begin{matrix}\begin{matrix}\rho_{2,1} & \rho_{3,1} & \rho_{6,1} & \rho_{3,2} & \rho_{3,6}\end{matrix}\\ \begin{pmatrix}20 & 40 & \mathbf{100} & 20 & 20\\ \mathbf{100} & 40 & 20 & 20 & 0\\ 40 & \mathbf{100} & 40 & 40 & 40\\ 20 & 40 & 20 & 20 & \mathbf{100}\\ 20 & 40 & 20 & \mathbf{100} & 20\end{pmatrix}\end{matrix}$$

Similarity = 100%

Categorize Objects: Principles

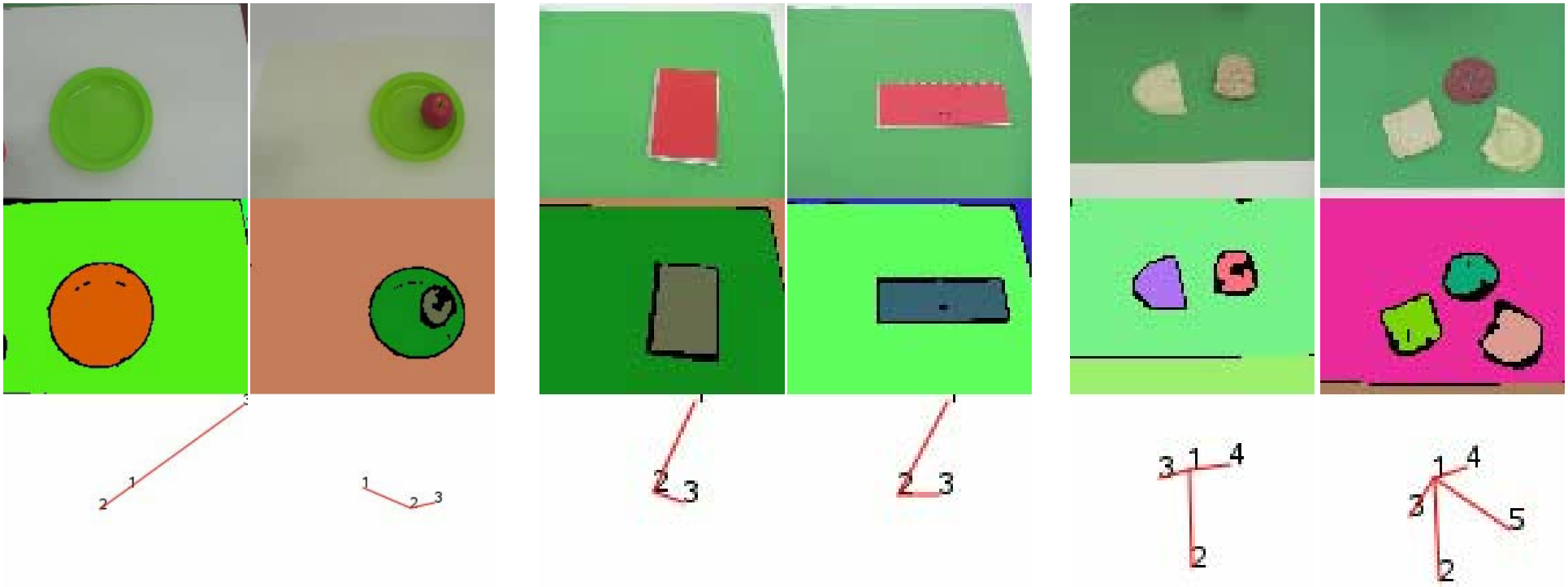
Algorithm – The algorithm can extract Semantic Similarity in scenes with different numbers of objects and different motions as well as orientations, etc.



The algorithm can categorize object with same roles in a manip.

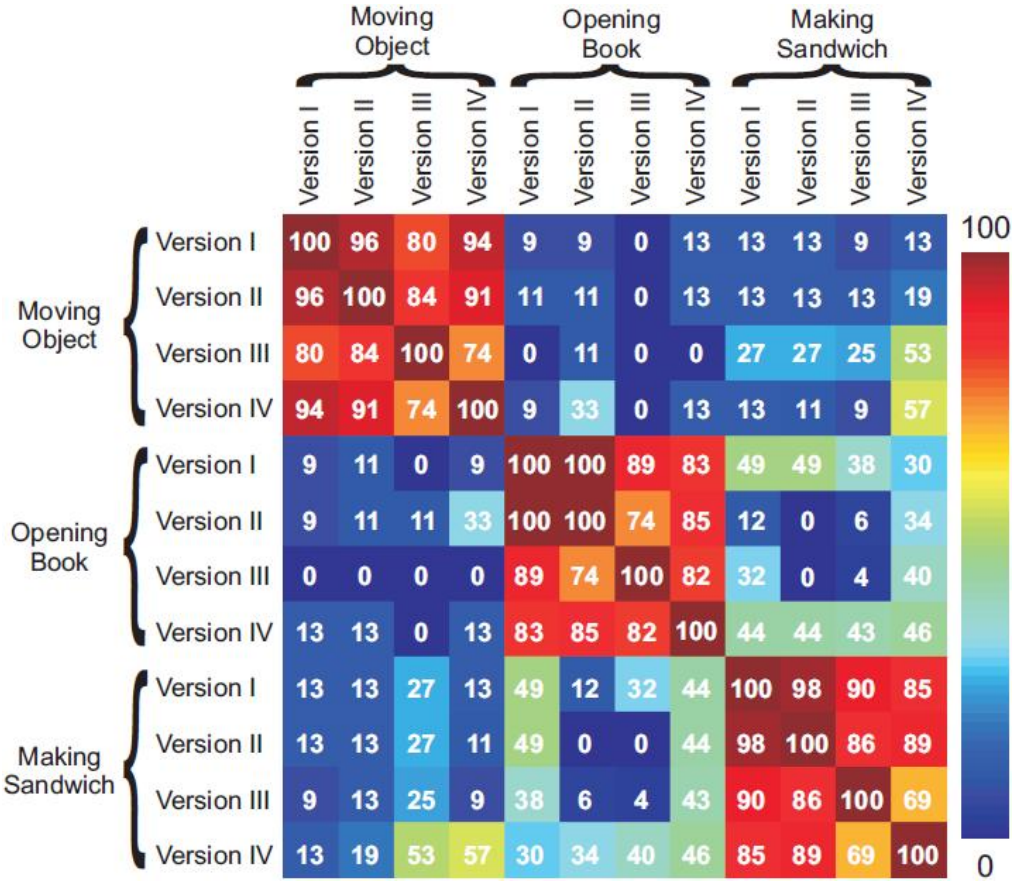
Analyse Manipulations

Action I: “Placing object”, “Opening Book” and “Making a sandwich”

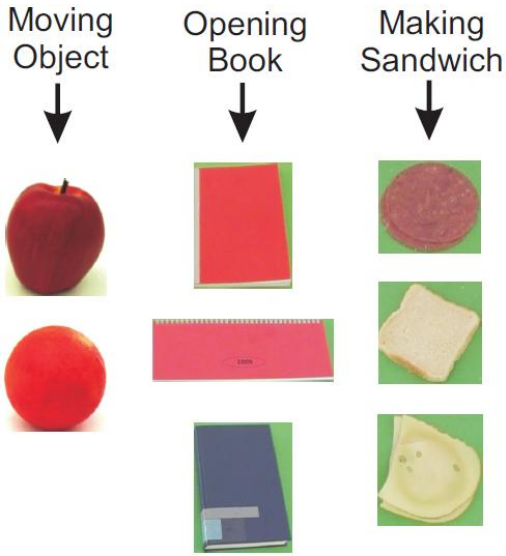


Analyse Manipulations

Action Classification

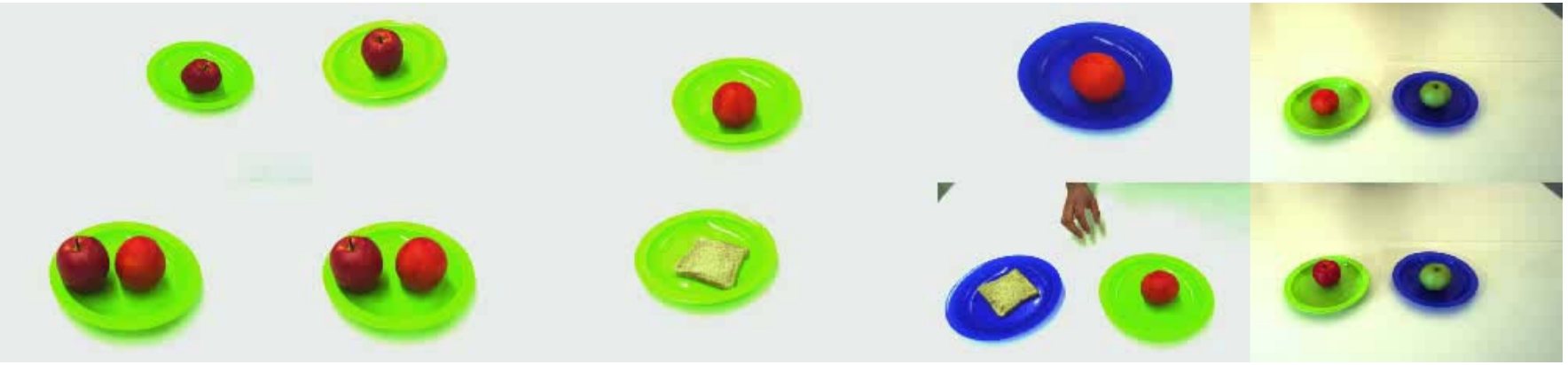


Object Categorization



Learning

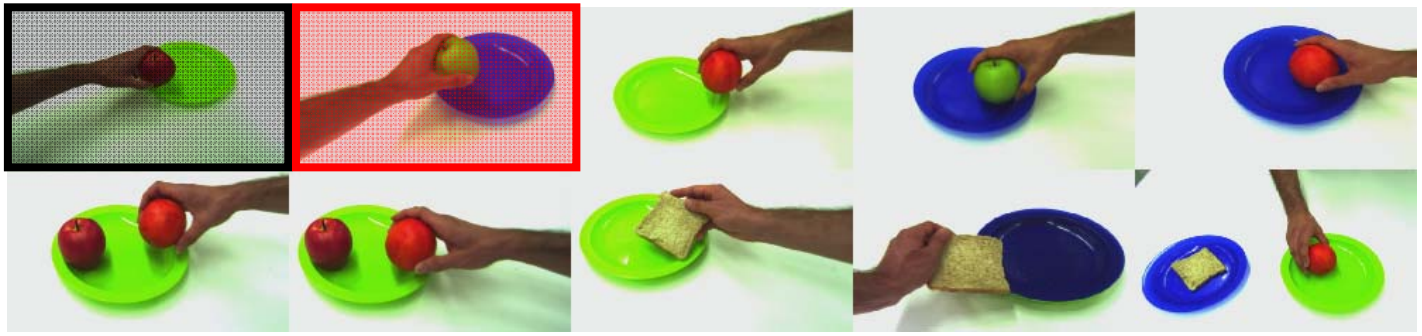
“Putting an Object on a Plate”



“Taking an Object from a Plate”

Learning: Exclusion of false positives

Learning



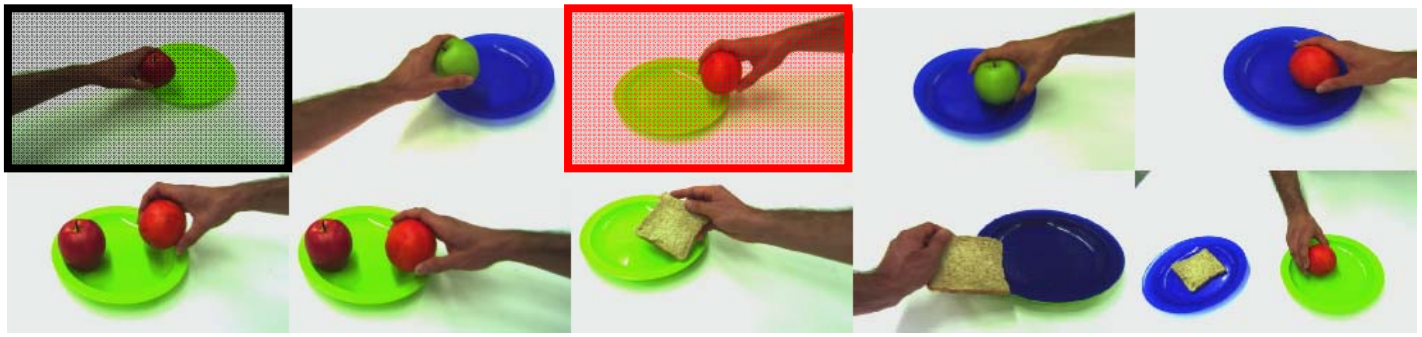
d/dt

$\rho_{2,1}$	I	L		100% →	$\omega_{2,1}$ =1
$\rho_{3,1}$	C	K		25% →	$\omega_{3,1}$ =1
$\rho_{5,1}$	C	J		100% →	$\omega_{5,1}$ =1
$\rho_{8,1}$	C	K	D	→	$\omega_{8,1}$ =1
$\rho_{3,2}$	A	F	L	75% →	$\omega_{3,2}$ =1
$\rho_{5,2}$	A	F	K D	100% →	$\omega_{5,2}$ =1
$\rho_{8,2}$	C	L	G	→	$\omega_{8,2}$ =1
$\rho_{5,3}$	C	K	D	100% →	$\omega_{5,3}$ =1
$\rho_{8,3}$	C	J		→	$\omega_{8,3}$ =1
$\rho_{8,5}$	C	K	D	→	$\omega_{8,5}$ =1

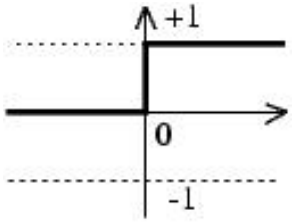
Repetition leads to the exclusion of spurious hence false action transitions (those with small weights).

Learning: Exclusion of false positives

Learning



$\rho_{2,1}$	I	L			100% → $\omega_{2,1}$ = 30
$\rho_{3,1}$	C	K			100% → $\omega_{3,1}$ = 1
$\rho_{5,1}$	C	J			100% → $\omega_{5,1}$ = 8
$\rho_{8,1}$	C	K	D		→ $\omega_{8,1}$ = 2
$\rho_{3,2}$	A	F	L		75% → $\omega_{3,2}$ = 3
$\rho_{5,2}$	A	F	K	D	100% → $\omega_{5,2}$ = 9
$\rho_{8,2}$	C	L	G		→ $\omega_{8,2}$ = 3
$\rho_{5,3}$	C	K	D		100% → $\omega_{5,3}$ = 30
$\rho_{8,3}$	C	J			→ $\omega_{8,3}$ = 4
$\rho_{8,5}$	C	K	D		→ $\omega_{8,5}$ = 2



Hard-limit transfer function

$\rho_{2,1}$	I	L		
$\rho_{3,1}$	C	K		
$\rho_{5,1}$	C	J		
$\rho_{3,2}$	A	F	L	
$\rho_{5,2}$	A	F	K	D
$\rho_{5,3}$	C	K	D	

Learned Model

Learning – Test Data

Take then put

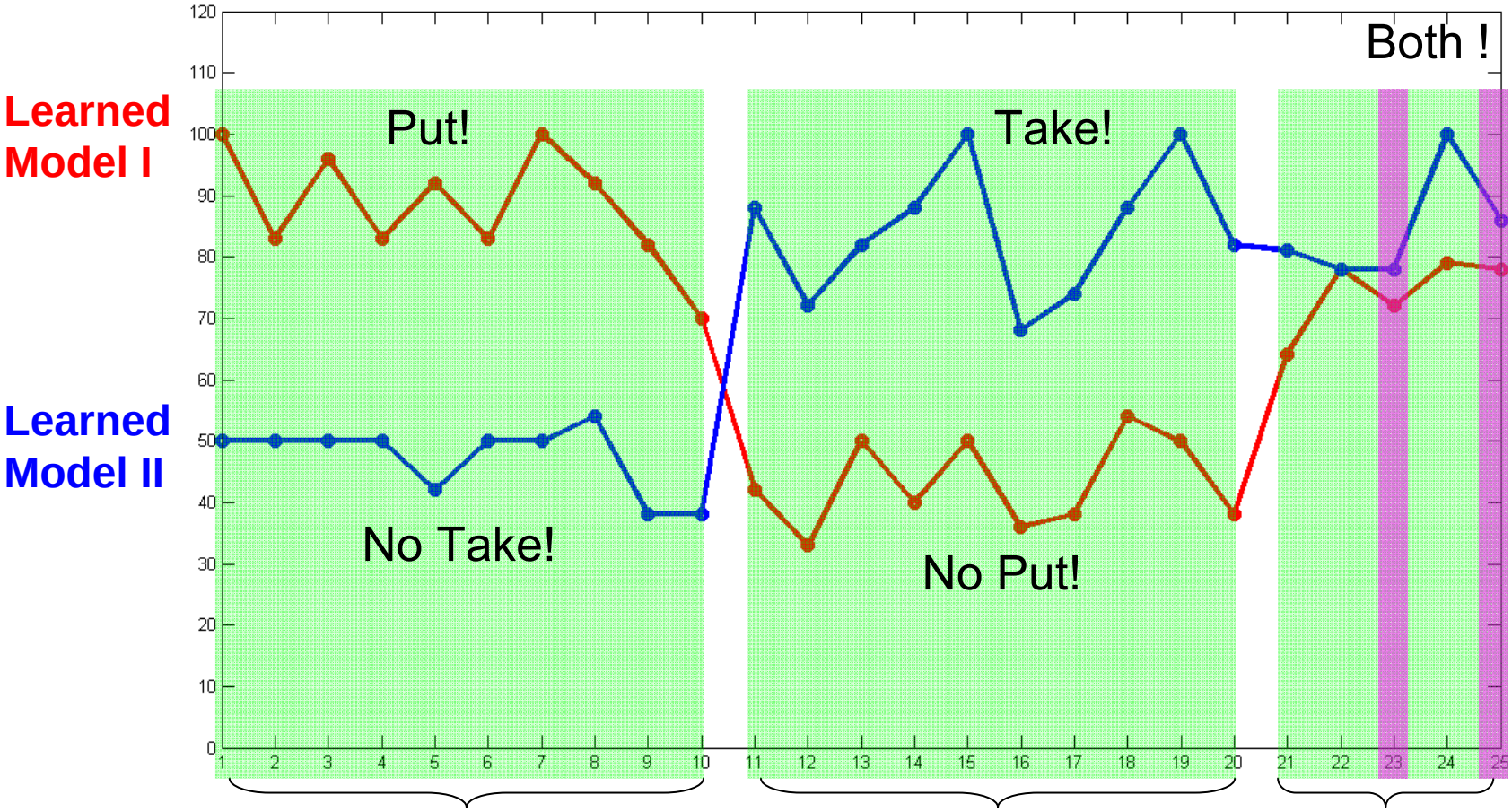
Put and take

Put then take other

Mixed Actions: Put and take ALWAYS occur here in some order

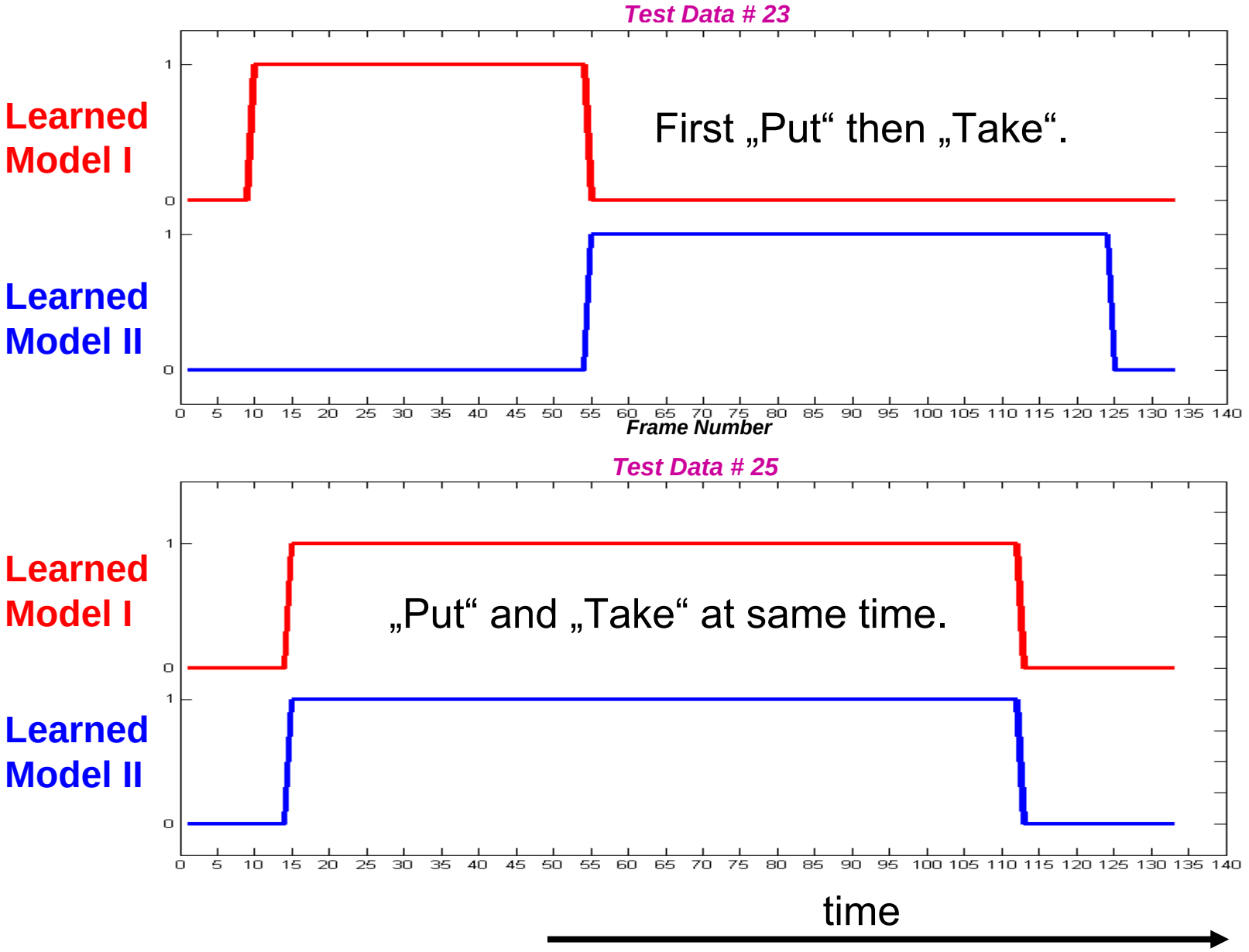
Learning: Results

General recognition



Learning: Results

Temporal recognition



Learning – Test Data

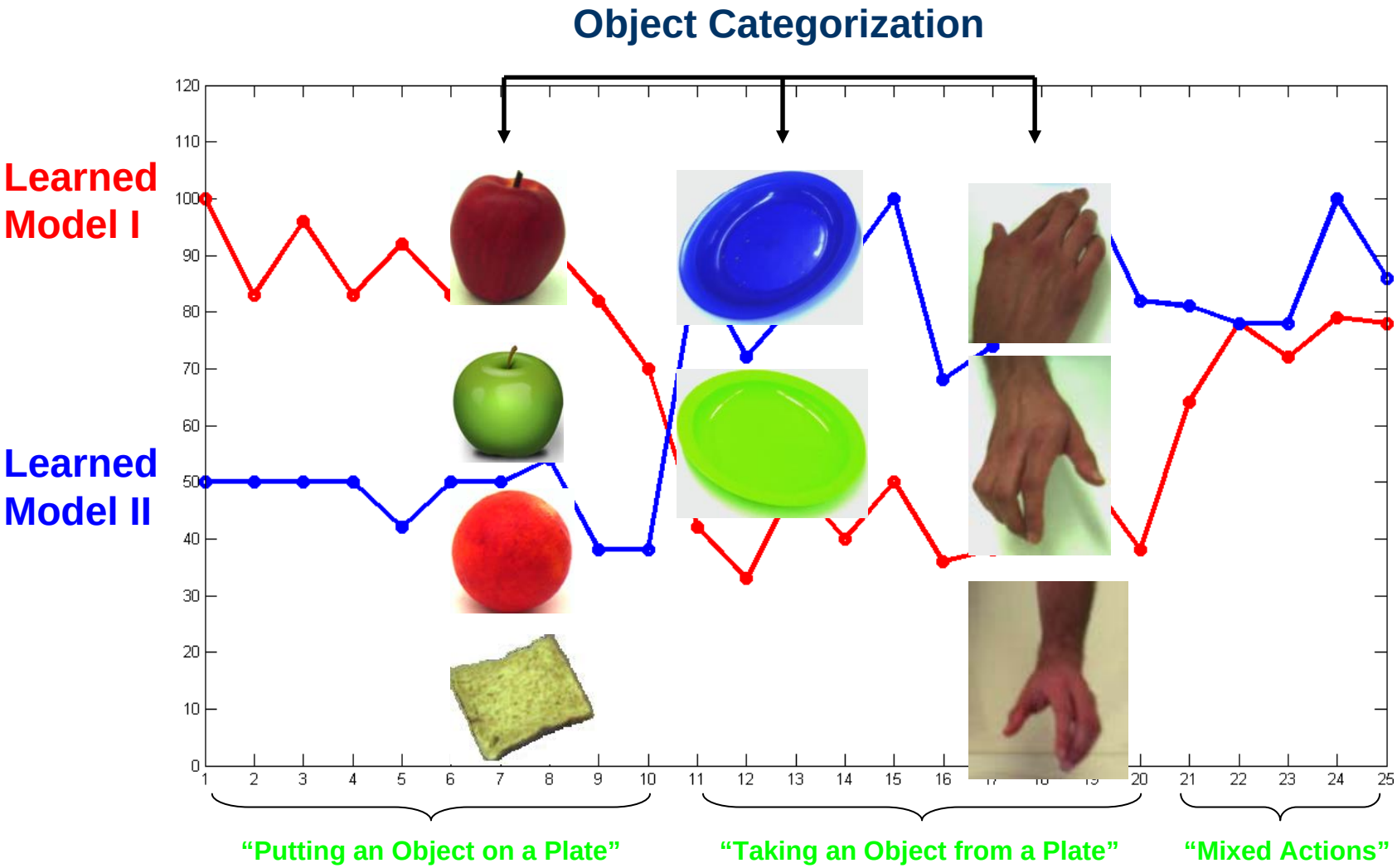
25

„Putting on“
and „taking from“
at same time

23

„Putting on“
and „taking from“
subsequently

Learning: Results: Object categorization



Model Free

Manipulation recognition using a counting algorithm.

Decomposible into chunks (anti-combi-explos!)

Learning the „meaning“ of a sequence of actions

Object Categorization as to their role in a manipulation.

Segments may(?) offer to hook on the object domain

(Some nice stereo and segmentation stuff

GPU-based video real time (mostly))


The people behind this work are:
Babette Dellen: Segmentation, tracking and stereo algorithms
Eren Aksoy: Graphs and Recognition
Alexey Abramov: All real time stuff
Johannes Dörr: Object models in the same framework (not discussed)