STIDS 2010

SEMANTIC TECHNOLOGY FOR
INTELLIGENCE, DEFENSE, AND SECURITY

Introducing Ontological Realism for Unsupervised Detection and Annotation of Operationally Significant Activity in Surveillance Videos

October 27-28, 2010
George Mason University, Fairfax, Virginia

W. Ceusters[1] , J. Corso[2], Y. Fu[2], M. Petropoulos[2], V. Krovi[3]

[1] *Ontology Research Group, NYS Center of Excellence in Bioinformatics & Life Sciences, 701 Ellicott Street, Buffalo NY*
[2] *School of Computer Science and Engineering, University at Buffalo*
[3] *Department of Mechanical and Aerospace Engineering, University at Buffalo*

New York State
Center of Excellence in
Bioinformatics & Life Sciences

The ISTARE Team

Jason Corso
Computer Science and Engineering

Computer Vision
Statistical Learning
Ontological Reasoning

ISTARE
Intelligent Spatiotemporal Activity Reasoning Engine

Raymond Fu
Computer Science and Engineering

Computer Vision
Manifold Learning

Werner Ceusters
Psychiatry
Ontology Research Group

Ontology
Referent Tracking

Michalis Petropoulos
Computer Science and Engineering

Databases Storage
Query Languages

Venkat Krovi
Mechanical and Aerospace Engineering

Robotics
Biomechanics
Articulated Motion

DARPA's Mind's Eye Program (1)

- **Purpose**: develop software for a smart camera, which is mountable on, f.i., man-portable UGVs and which exhibits capabilities necessary to perform surveillance in operational missions.
- **Capabilities requested**:
  - recognize the primitive actions that take place between objects in the visual input, with a particular emphasis on actions that are relevant in typical operational scenarios (e.g., vehicle APPROACHES checkpoint; person EXITS building).

DARPA's Mind's Eye Program (2)

- **Capabilities requested (continued):**
  - learning and cross-scene application of invariant spatio-temporal patterns,
  - issuing alerts to activities of interest,
  - performing interpolation to fill in likely explanations for gaps in the perceptual experience,
  - explaining its reasoning by displaying relevant video segments for what has been observed, and by generating visualizations for what is hypothesized.

Actions of interest

approach	carry	dig	fall	give	hit	lift
arrive	catch	drop	flee	go	hold	move
attach	chase	enter	fly	hand	kick	open
bounce	close	exchange	follow	haul	jump	pass
bury	collide	exit	get	have	leave	pick up

push	run	touch
put down	snatch	turn
raise	stop	walk

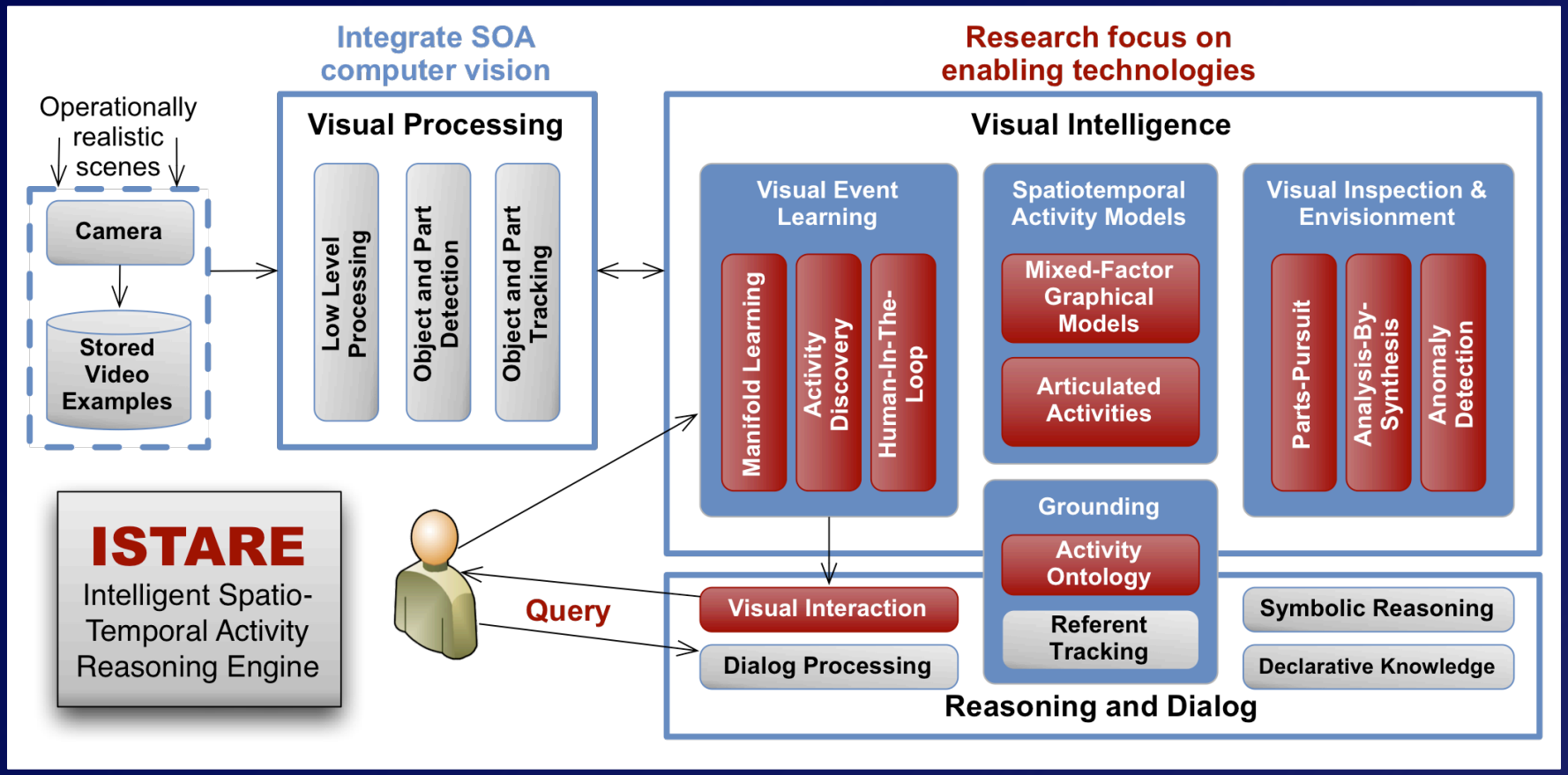
pass	receive	take
pick up	replace	throw

Run - Take

Drop - Lift

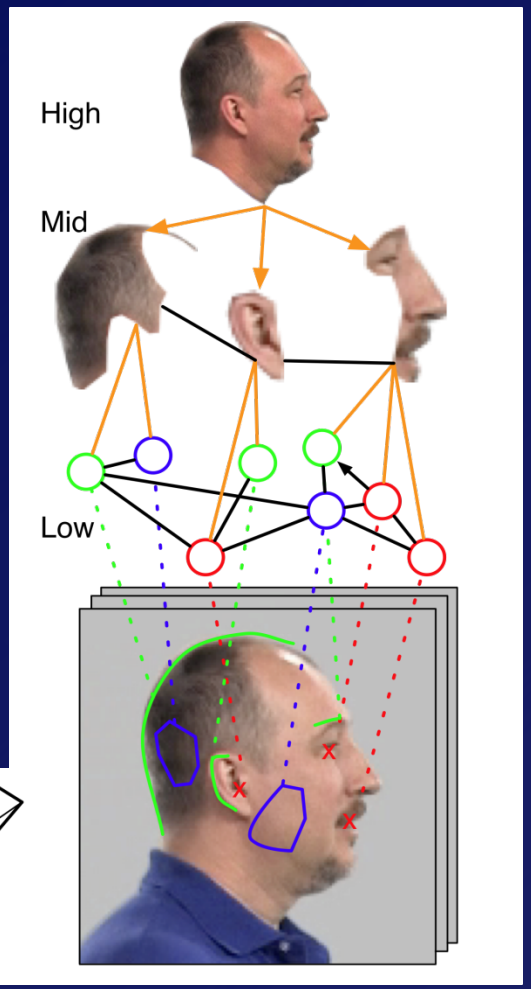
Filling gaps

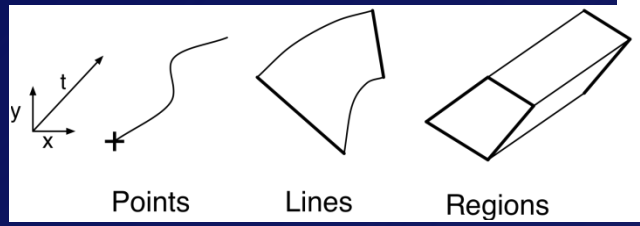
System overview



Representation strategy

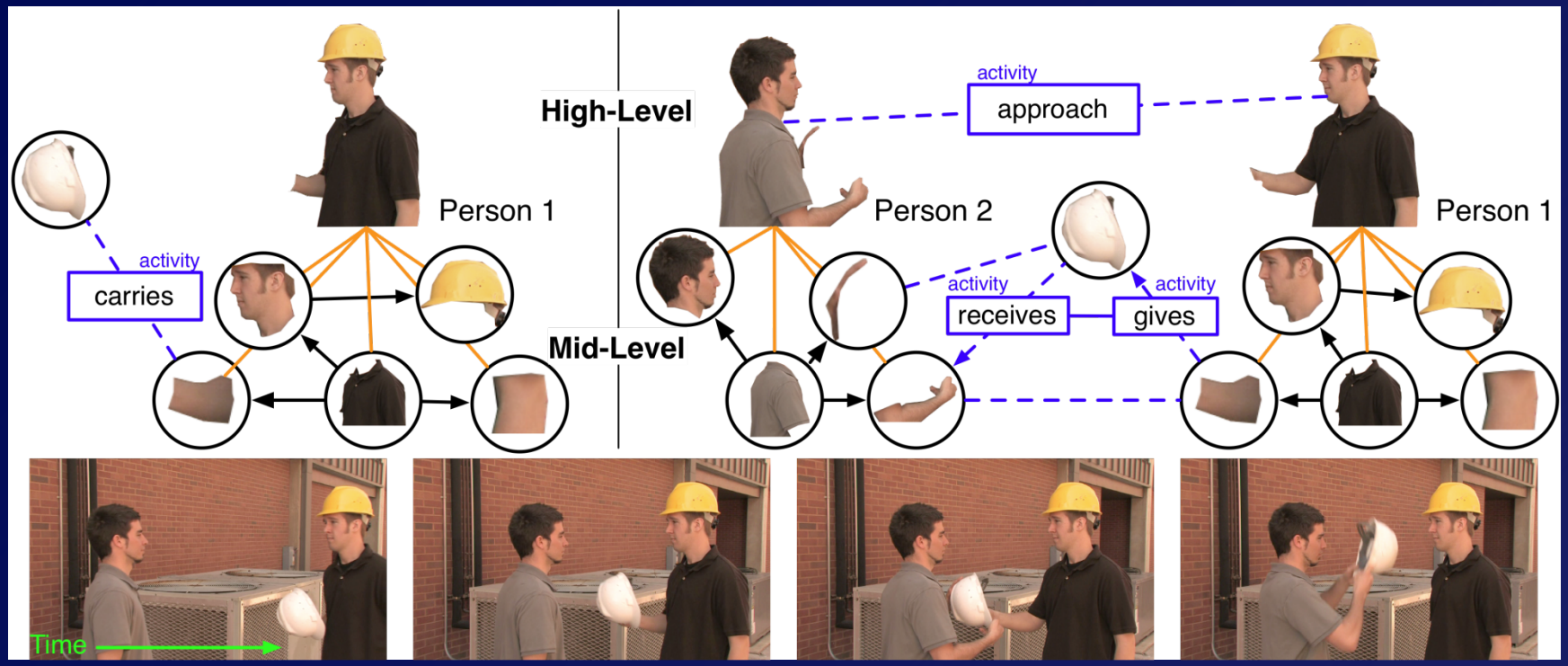
- Hierarchical and generative representation.
- Allows for reuse of parts at the low- and mid-levels.
- Each object is a rooted graph with a mixture of directed and undirected edges, which govern the nature of the connections.
- At the low-level, atoms subtend spatiotemporal lines, planes, and regions over which we learn non-linear manifolds.





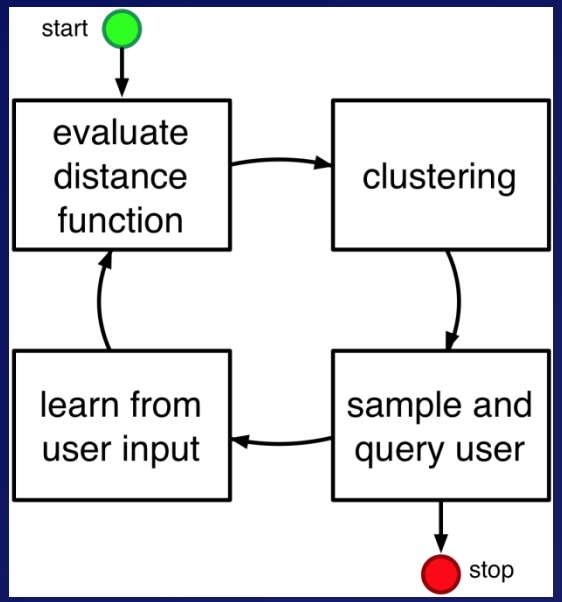
Representation of activities

- Activities are detected as changes in conglomerates of these object graphs, which have connections at both the mid- (part) and the high-level (object).



Learning with human in the loop

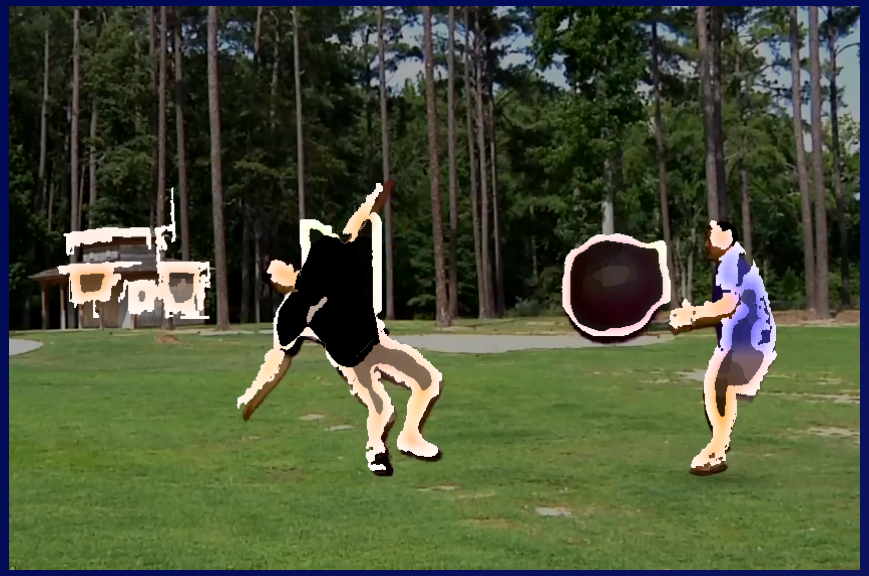
- For learning the structure at the mid-to-high levels.
  - Proceed with an active semi-supervised learning schema to iteratively refine a distance metric.
  - At mid-level: clustering is the game.
    - Iteratively seek grouping constraints.
    - Interactively link clusters to ***ontology***.
    - Working with both spectral and hierarchical clustering.
  - At high-level: get feedback from the user based on our ***ontological*** grounding: ask Questions, such as "is the person in this clip giving the object to the other person?"



ISTARE Ontology

- Roles:
  - Learning: it will help guide a learning algorithm to remain in plausible configurations.
  - Inference: it will support reasoning of plausible explanations of objects and activities in existing and missing parts of the signal.
- Components:
  - L1 ⇔ L1:
    - How humans interact with objects and other humans in various scenarios.
    - How motions of object-parts contribute to full object motion.
  - L1 ⇔ L3:
    - How manifolds in the video correspond to entities videotaped.
  - L1 ⇔ L2 ⇔ L3:
    - How analysts interpret videos and corresponding reality.

L1 – L3



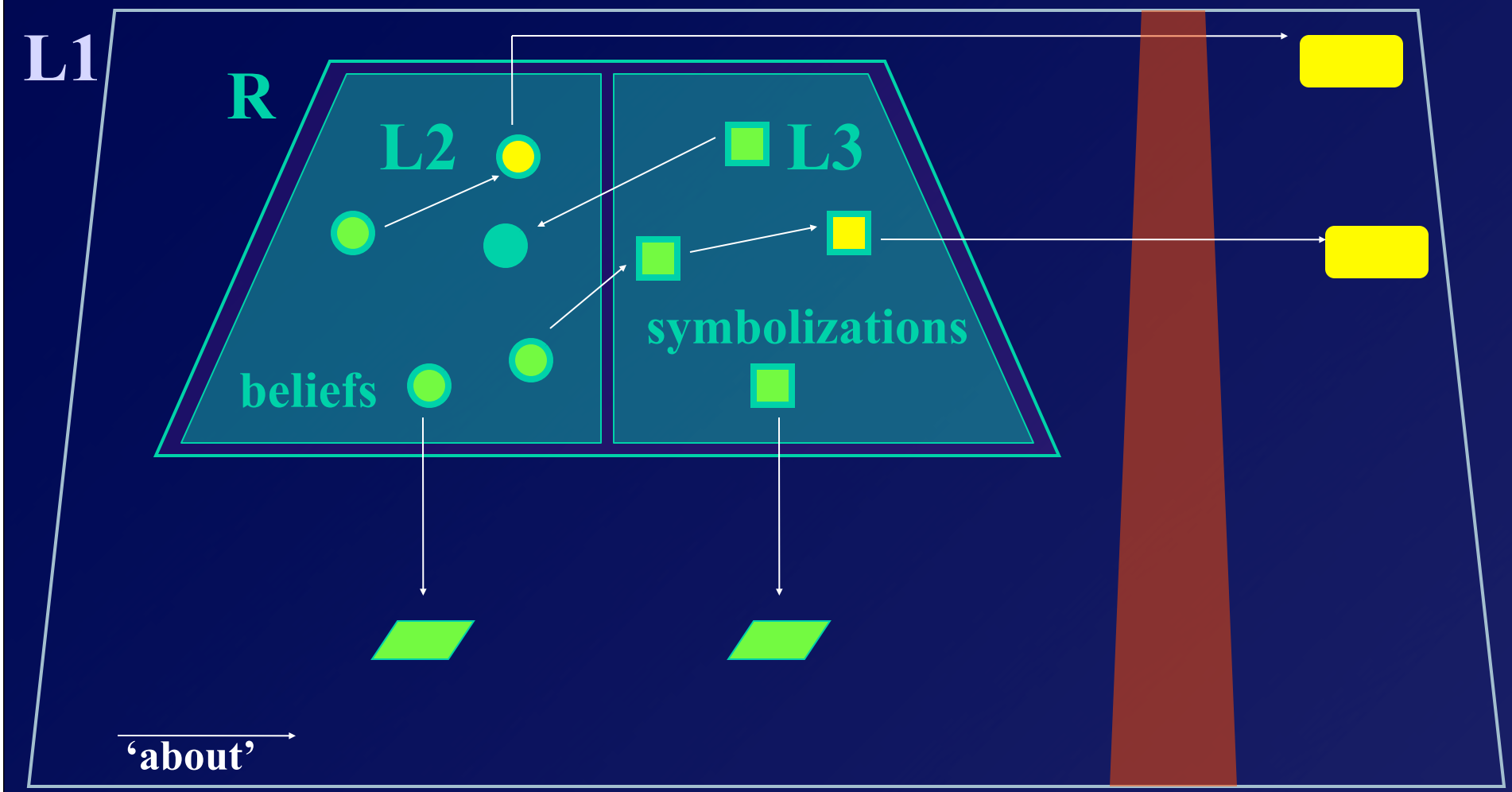
The basis: Ontological Realism

Basic axioms of realism:

1. There is an external reality which is 'objectively' the way it is;
2. That reality is accessible to us;
3. We build in our brains cognitive representations of reality;
4. We communicate with others about what is there, and what we believe there is there.

Smith B, Kusnierczyk W, Schober D, Ceusters W. Towards a Reference Terminology for Ontology Research and Development in the Biomedical Domain. Proceedings of KR-MED 2006, Biomedical Ontology in Action, November 8, 2006, Baltimore MD, USA

Three levels of reality, two sorts of representations



Main distinctions in BFO

some *continuant* universal

some *occurrent* universal

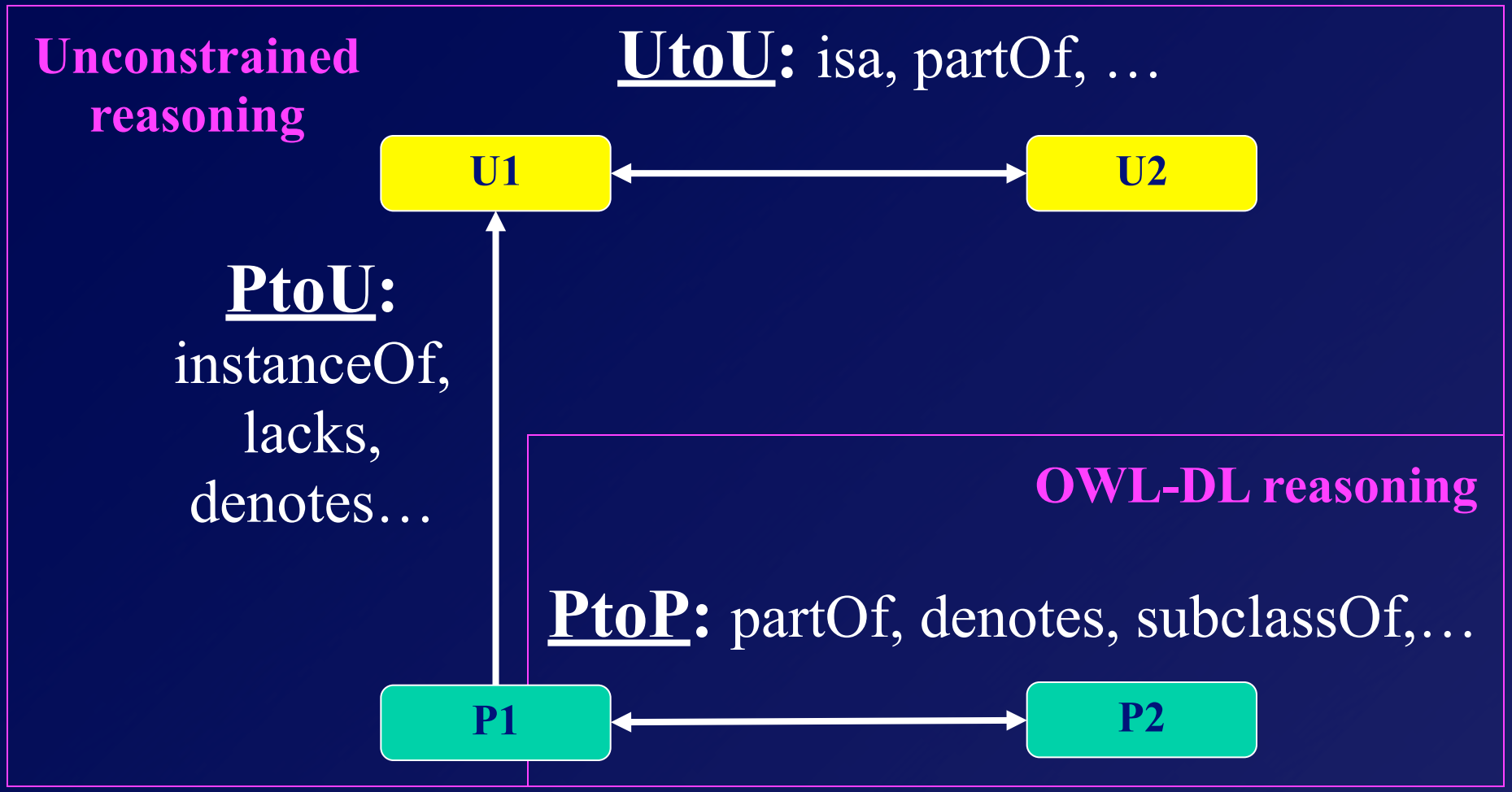
instanceOf <u>at t</u>

instanceOf

some *continuant* particular

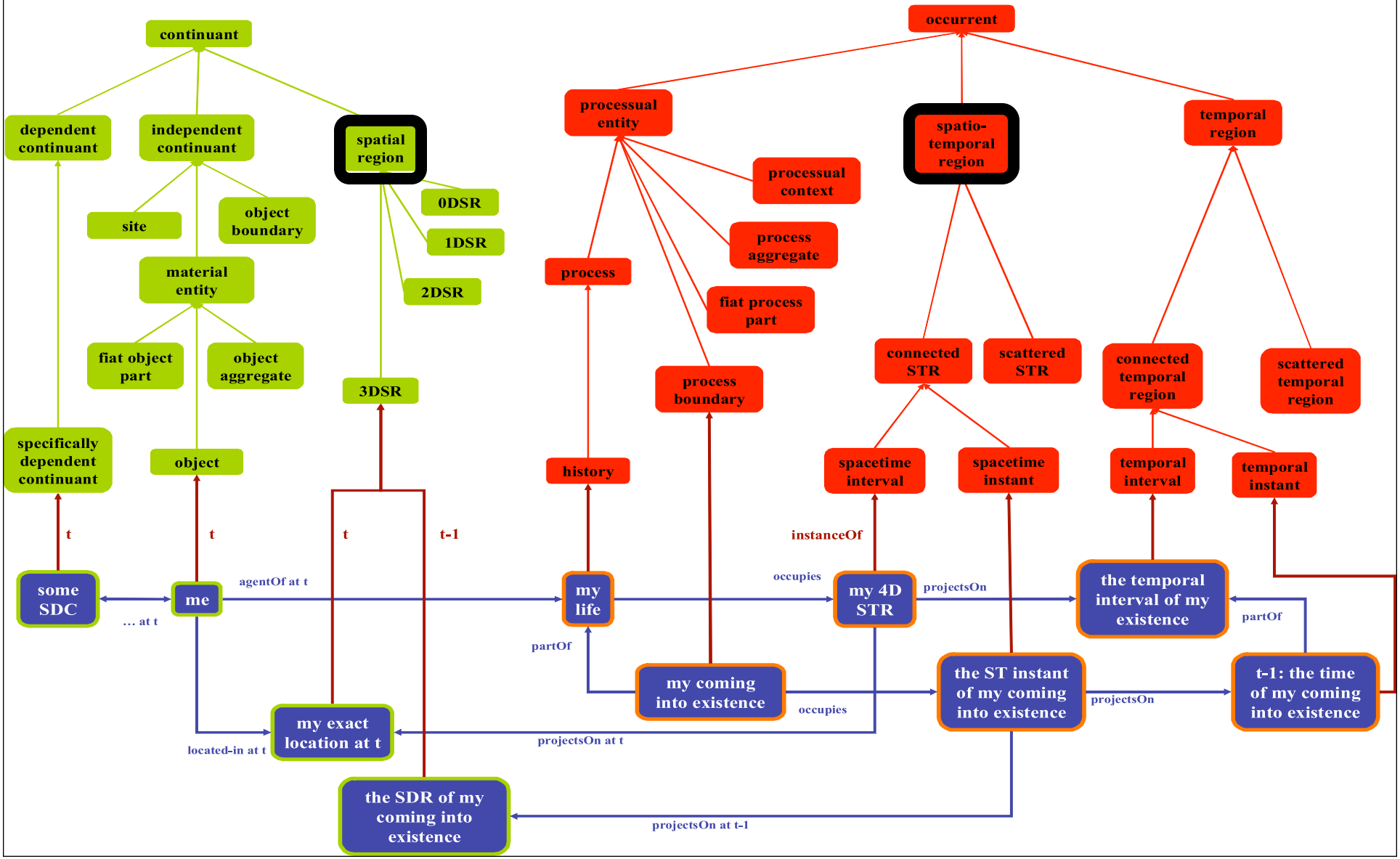
some *occurrent* particular

Sorts of relations



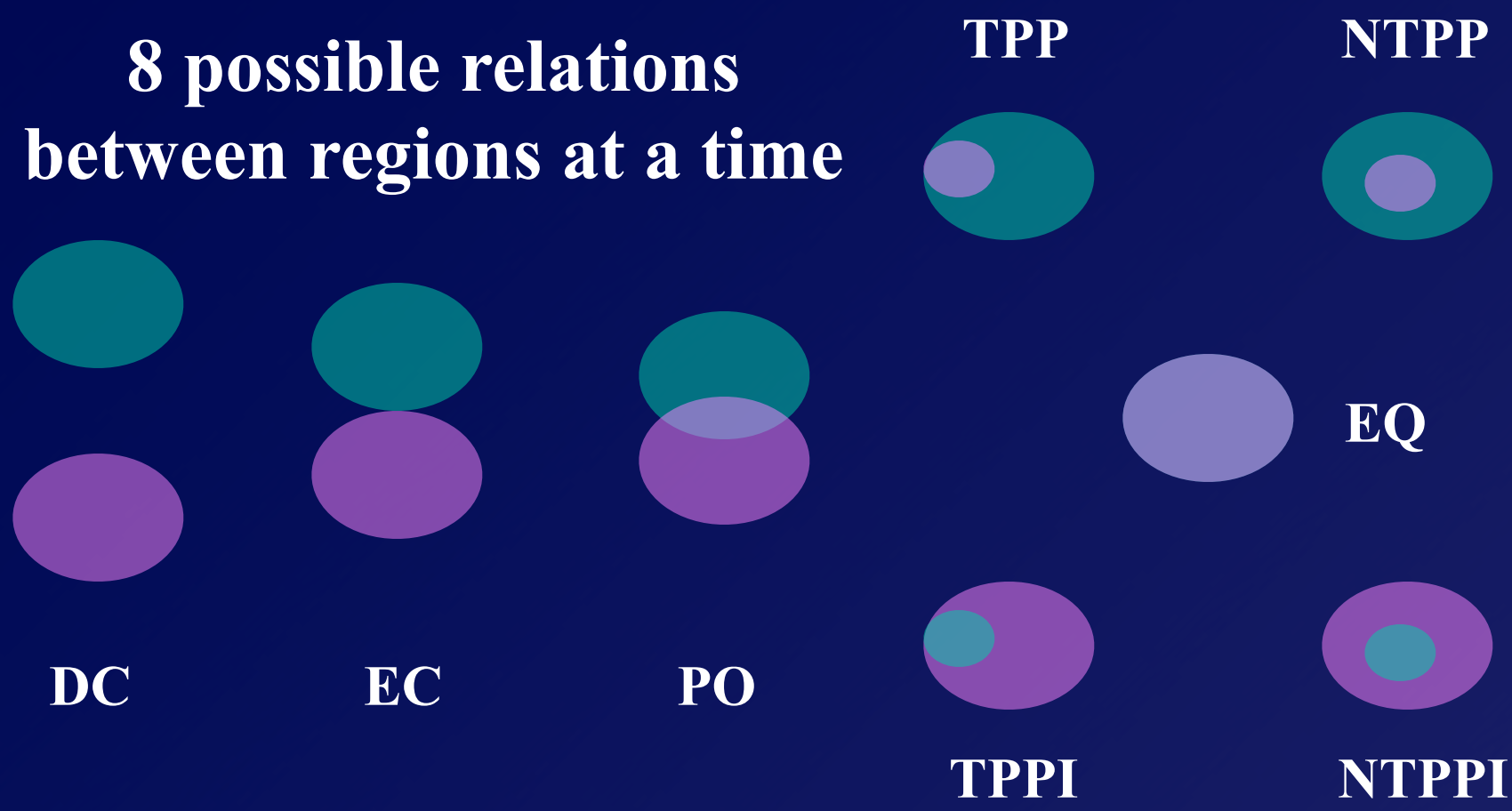
New York State
Center of Excellence in
Bioinformatics & Life Science

Relevant First-Order Distinctions



Region Connection Calculus (RCC8)

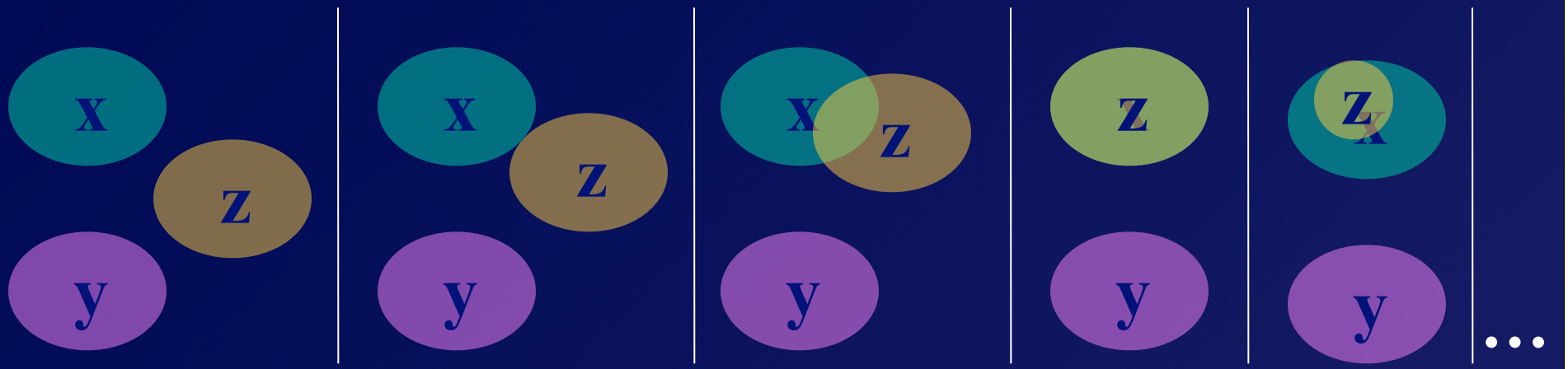
8 possible relations between regions at a time



Randell, D., Cui, Z., Cohn, A.: A Spatial Logic based on Regions and Connection.
In: Proceedings of the International Conference on Knowledge Representation and Reasoning, pp. 165–176 (1992)

RCC8 reasoning

- $rel_1(x,y,t) \wedge rel_2(y,z,t) \rightarrow rel_3(x,z,t)$?
- e.g. $DC(x,y,t) \wedge DC(y,z,t)$

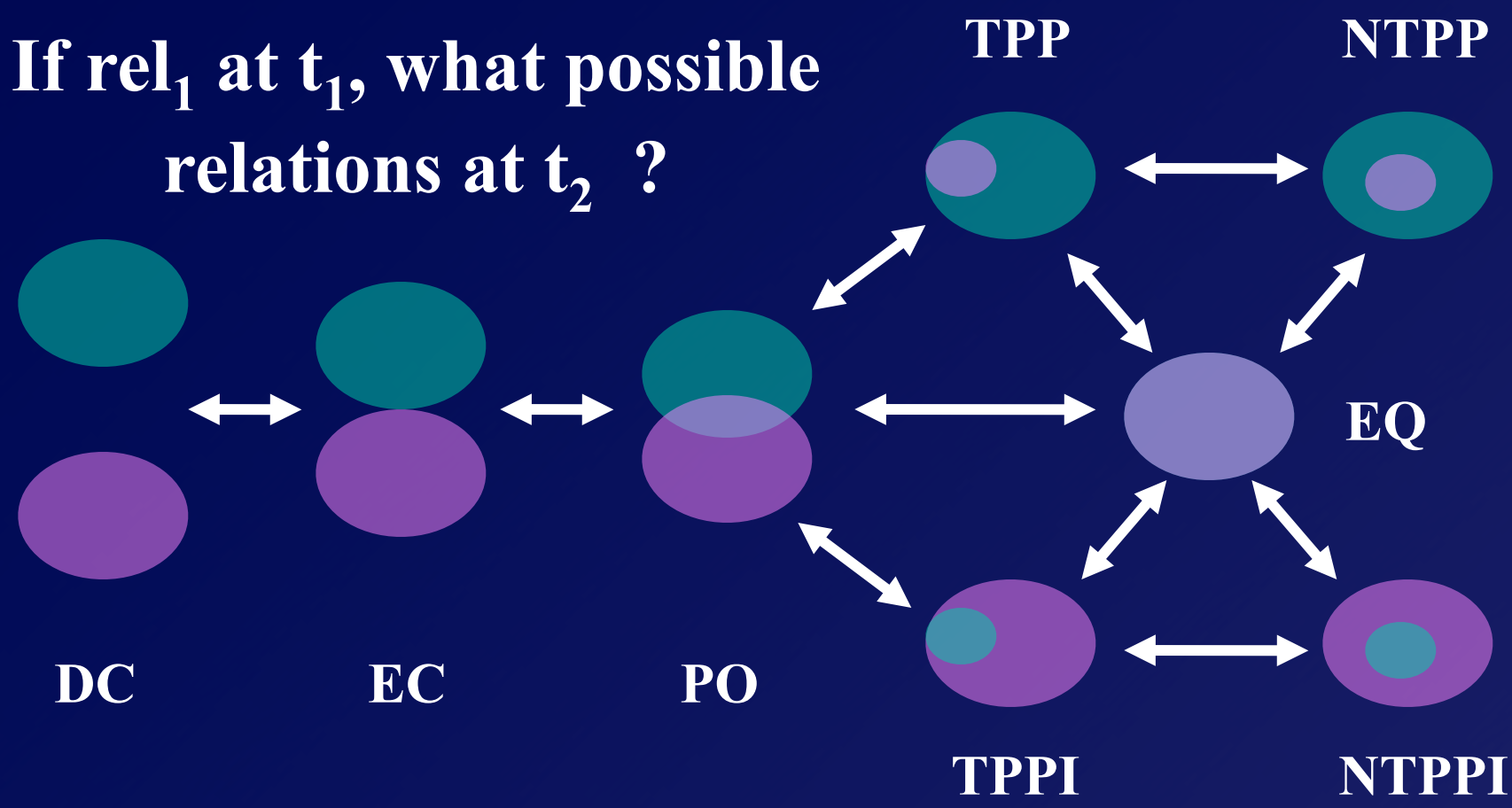


- maintained in tables

Randell, D., Cui, Z., Cohn, A.: A Spatial Logic based on Regions and Connection.
In: Proceedings of the International Conference on Knowledge Representation and Reasoning, pp. 165–176 (1992)

RCC8: conceptual neighborhood

If rel_1 at t_1, what possible relations at t_2 ?



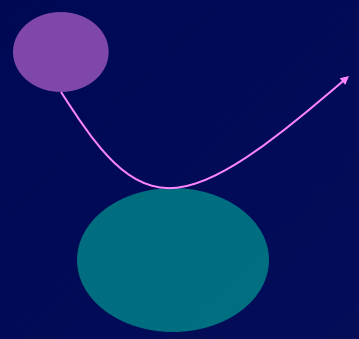
Randell, D., Cui, Z., Cohn, A.: A Spatial Logic based on Regions and Connection.
In: Proceedings of the International Conference on Knowledge Representation and Reasoning, pp. 165–176 (1992)

Basic '*Motion Classes*'

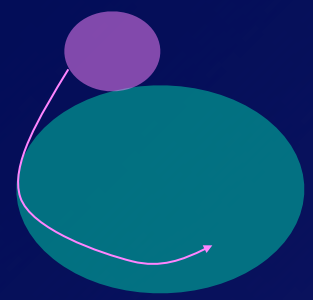
		Ends							
		DC	EC	PO	TPP	NTPP	EQ	TPPI	NTPPI
Starts	DC	External	Hit	Reach					
	EC	Split	Peripheral						
	PO			Leave or Reach					
	TPP				Internal		Expand		
	NTPP								
	EQ						Internal		
	TPPI				Shrink			Internal	
	NTPPI								

Zina Ibrahim, and Ahmed Y. Tawfik, An Abstract Theory and Ontology of Motion Based on the Regions Connection Calculus, Symposium of Abstraction, Reformulation and Approximation (SARA 2007), LNAI, Springer, 2007.

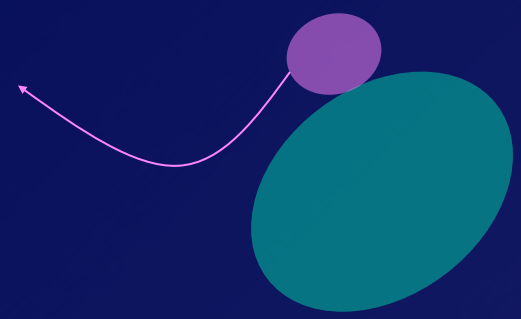
Compound motion classes



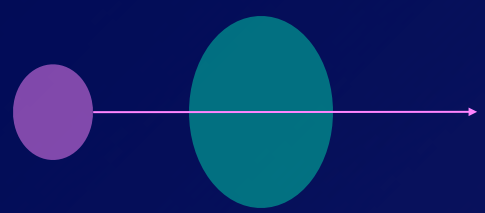
hit-split



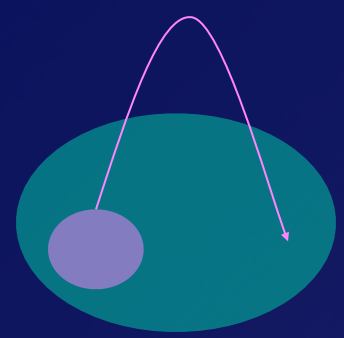
peripheral-reach



peripheral-leave



reach-leave

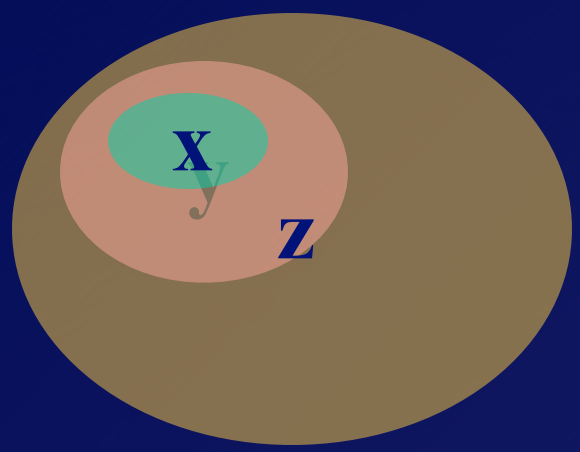


leave-reach

Zina Ibrahim, and Ahmed Y. Tawfik, An Abstract Theory and Ontology of Motion Based on the Regions Connection Calculus, Symposium of Abstraction, Reformulation and Approximation (SARA 2007), LNAI, Springer, 2007.

Reasoning with motion classes

- $mc_1(x,y,t) \wedge mc_2(y,z,t) \rightarrow mc_3(x,z,t)$?
- e.g. $leave(x,y,t) \wedge leave(y,z,t)$

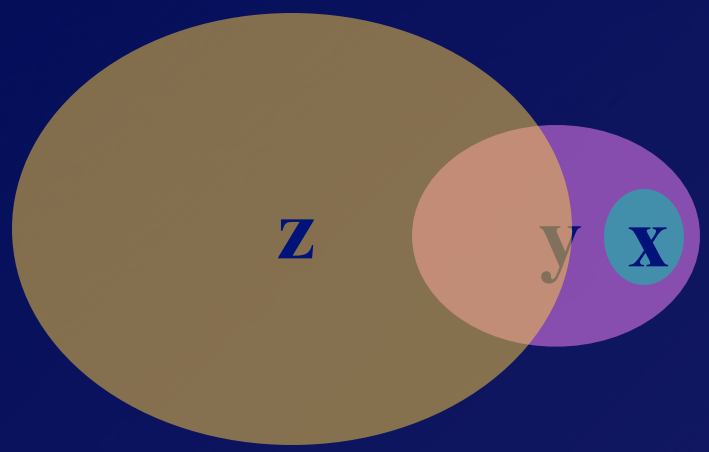


internal

Zina Ibrahim, and Ahmed Y. Tawfik, An Abstract Theory and Ontology of Motion Based on the Regions Connection Calculus, Symposium of Abstraction, Reformulation and Approximation (SARA 2007), LNAI, Springer, 2007.

Reasoning with motion classes

- $mc_1(x,y,t) \wedge mc_2(y,z,t) \rightarrow mc_3(x,z,t)$?
- e.g. $leave(x,y,t) \wedge leave(y,z,t)$

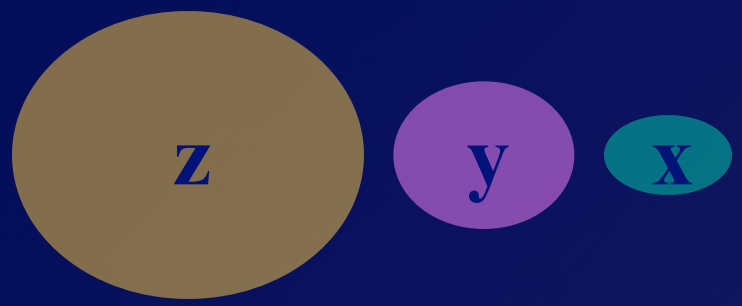


external

Zina Ibrahim, and Ahmed Y. Tawfik, An Abstract Theory and Ontology of Motion Based on the Regions Connection Calculus, Symposium of Abstraction, Reformulation and Approximation (SARA 2007), LNAI, Springer, 2007.

Reasoning with motion classes

- $mc_1(x,y,t) \wedge mc_2(y,z,t) \rightarrow mc_3(x,z,t)$?
- e.g. $leave(x,y,t) \wedge leave(y,z,t)$



- all possibilities also in tables

Zina Ibrahim, and Ahmed Y. Tawfik, An Abstract Theory and Ontology of Motion Based on the Regions Connection Calculus, Symposium of Abstraction, Reformulation and Approximation (SARA 2007), LNAI, Springer, 2007.

RCC8/MC14 and Ontological Realism

- In ontological realism:
  - regions don't move
  - material entities are located in regions
  - while material entities move or shrink/expand:
    - they are located at each t in a different region
    - each such region is part of the region formed by all the regions visited, thus constituting a path
    - …
- An unambiguous mapping is possible

RCC8/MC14 and action verbs

‘approach’

RCC8/MC14 and action verbs

'approach'

- Invariant:
  - ***shrink*** of the region between the entities involved in an approach

RCC8/MC14 and action verbs

approach	carry	dig	fall	give	hit	lift	push	run	touch
arrive	catch	drop	flee	go	hold	move	put down	snatch	turn
attach	chase	enter	fly	hand	kick	open	raise	stop	walk
bounce	close	exchange	follow	haul	jump	pass	receive	take	
bury	collide	exit	get	have	leave	pick up	replace	throw	

- all can be expressed in terms of mc14 (with the addition of direction and some other features)
- from mc to the verbs: requires additional information on the nature of the entities involved
  - to be encoded in the ontology

Action verbs and Ontological Realism

- Many caveats:
  - the way matters are ***expressed*** in natural language does not correspond faithfully with the way matters ***are***

'approach'

x orbiting around y

x taking distance from y ?

x approaching y ?

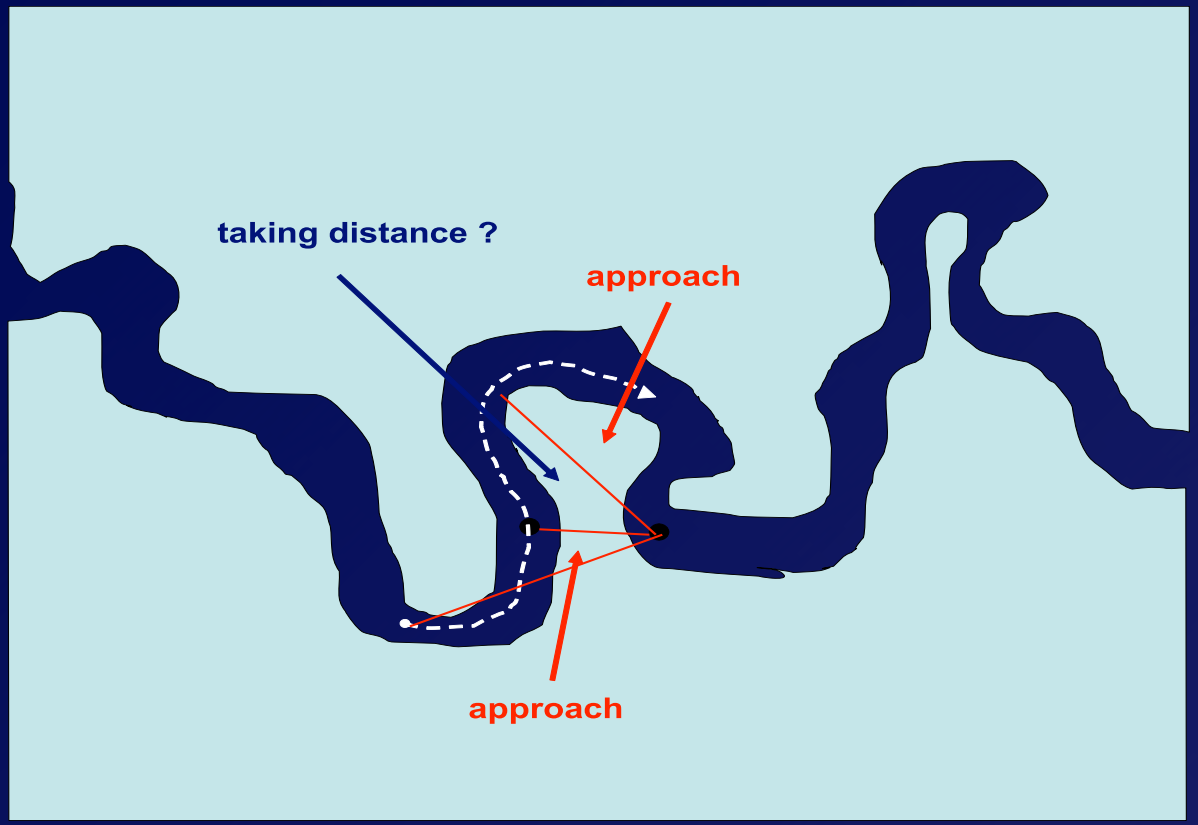
x taking distance from y ?

→ x's process didn't change

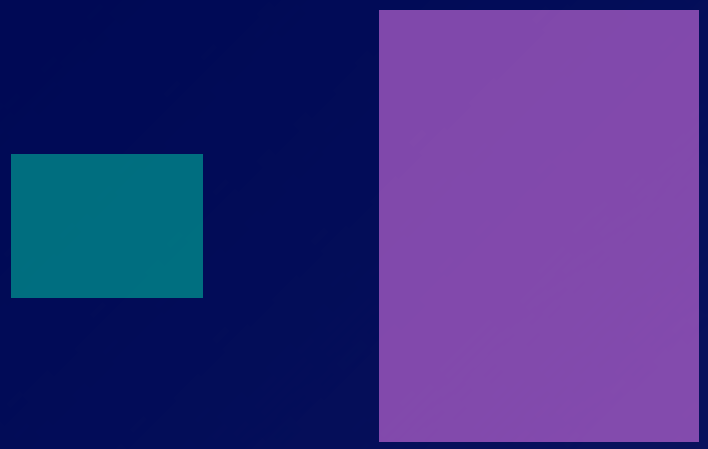
→ 'to approach' is a verb, but it does not represent a process, rather implies a process.

Action verbs and Ontological Realism

- Approaching following a forced path

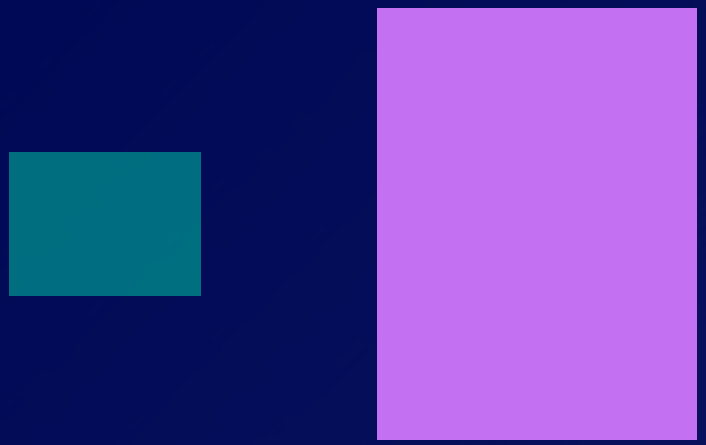


RCC8/MC14 & video as 2D+T representation of 3D+T



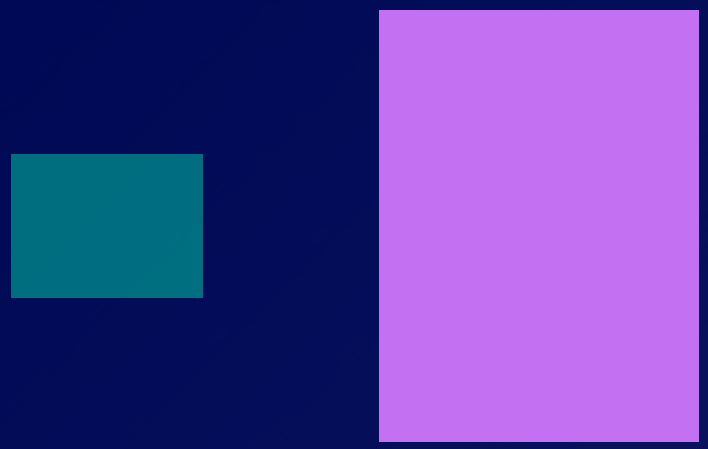
man entering building: the first-order view

RCC8/MC14 & video as 2D+T representation of 3D+T



man entering building: the video view

RCC8/MC14 & video as 2D+T representation of 3D+T



egg crashing on wall: the video view

- Requires additional mapping from the motion of manifolds in the video to the corresponding motion of the corresponding entities in reality

ISTARE Ontology design principles

- Main objective:
  - Being pragmatic, yet adhering to the principles of Ontological Realism.
- Each representational unit (RU) denotes either
  - a *universal:*
    - Human Being, Motion;
  - a *fiat class*, i.e, a portion of reality demarcated by human fiat, and relevant for the goals of ISTARE:
    - Canonically-limbed human being,
    - High five.

A partly worked out example

Human anatomy (L1)

- c1 member-of Canonically-Limbed Human Being at t, then:
  - c1 has-part c2 at t
  - c2 instance-of Left Arm at t
  - c1 has-part c3 at t
  - c3 instance-of Right Arm at t
  - c1 has-part c4 at t
  - c4 instance-of Left Lower Limb at t
  - c1 has-part c5 at t
  - c5 instance-of Right Lower Limb at t
  - c1 has-part c6 at t
  - c6 instance-of Head at t
  - c1 has-part c7 at t
  - c7 instance-of Torso at t
  - c2 adjacent-to c7 at t
  - c3 adjacent-to c7 at t
  - c4 adjacent-to c7 at t
  - c5 adjacent-to c7 at t
  - c6 adjacent-to c7 at t
  - …

Human physiology (L1)

- c1 member-of Canonically-Limbed Human Being at t, then:
  - sdc1 inheres-in c1 at t
  - sdc1 instance-of Disposition-to-Walk at t
  - sdc2 inheres-in c1 at t
  - sdc2 instance-of Disposition-to-Run at t
  - …

approach	carry	dig	fall	give	hit	lift	push	run	touch
arrive	catch	drop	flee	go	hold	move	put down	snatch	turn
attach	chase	enter	fly	hand	kick	open	raise	stop	walk
bounce	close	exchange	follow	haul	jump	pass	receive	take	
bury	collide	exit	get	have	leave	pick up	replace	throw	

impossible

under certain circumstances

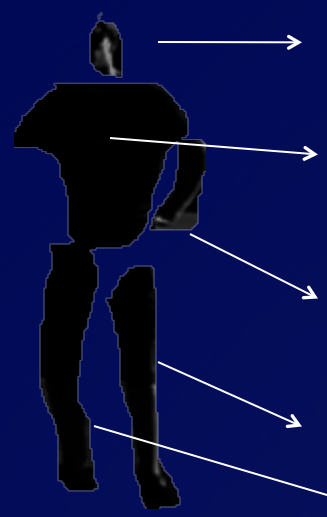
Human physiology (L1)

- o1 member-of Canonical-Human-Walking, then:
  - o1 realization-of sdc1
  - sdc1 instance-of Disposition-to-Walk at t
  - sdc1 inheres-in c1 at t
  - c1 instance-of Canonically-Limbed Human Being at t
  - o1 has-agent c1 at t
  - o1 has-part o2
  - o2 instance-of Walking Leg Motion
  - o2 has-agent c2 at t
  - c2 part-of c1 at t
  - c2 instance-of Left Lower Limb at t
  - o3 instance-of Walking Leg Motion
  - o3 has-agent c3 at t
  - c3 part-of c1 at t
  - c3 instance-of Right Lower Limb at t
  - c1 located-in r1 at t0
  - t0 earlier t
  - c1 located-in r2 at t1
  - t earlier t1
  - …

But: elliptical work-out, walking in circle, …

Human anatomy applied to video frames

- Detection and grounding of manifolds:



m1 denotes c6 at t ?

m2 denotes c7 at t ?

m3 denotes c2 at t ?

m4 denotes c4 at t ?

m5 denotes c5 at t ?

L3

L3

Elements of ontology-based reasoning

- Projection of RCC and MCC in L3 to portions of reality in L1:
  - EC → adjacent-to
  - shrink → shrinking
    → moving away from camera
  - hit → approach in front or behind object
  - hit < shrink → 'shrinking' object passed behind
  - …
- Human in the loop

Work in progress

- Finished:
  - Feasibility assessment by team
  - Core ontological principles
- Started:
  - Population of ontology
  - Documentation of invariants
- Next steps:
  - annotation of ‘verbs’ in terms of RUs from the ontology
  - development of reasoning engine
  - integration with manifold detection and tracking